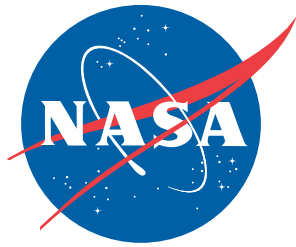


NASA/CP-2012-217326



Selected Papers Presented at MODSIM World 2011 Conference & Expo

Compiled by

*Thomas E. Pinelli
Langley Research Center, Hampton, Virginia*

*Leanna S. Bullock
Alliant Techsystems, Inc., Langley Research Center, Hampton, Virginia*

March 2012

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

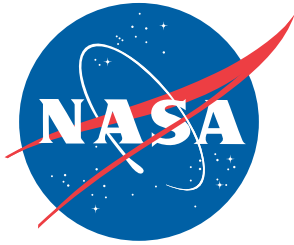
- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.
- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA STI Help Desk at 443-757-5803
- Phone the NASA STI Help Desk at 443-757-5802
- Write to:
NASA STI Help Desk
NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320

NASA/CP-2012-217326



Selected Papers Presented at MODSIM World 2011 Conference & Expo

Compiled by

*Thomas E. Pinelli
Langley Research Center, Hampton, Virginia*

*Leanna S. Bullock
Alliant Techsystems, Inc., Langley Research Center, Hampton, Virginia*

Proceedings of a conference sponsored by the
National Aeronautics and Space Administration
and held in Virginia Beach, Virginia
October 11–14, 2011

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

March 2012

The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320
443-757-5802

Preface

Selected papers from MODSIM World 2011 Conference & Expo are contained in this NASA Conference Publication (CP). MODSIM World 2011 was held in Virginia Beach, Virginia, October 11–14, 2011. The theme of the 2011 conference & expo was “Overcoming Critical Global Challenges with Modeling & Simulation”. The conference program consisted of five technical tracks – Defense, Homeland Security & First Responders; Education; Health & Medicine; The Human Dimension; and Serious Games & Virtual Worlds. As a condition for inclusion in the conference proceedings, the first author was responsible for securing/obtaining all permissions associated with the general release and public availability of the paper/presentation. Further, the first authors also had to grant NASA the right to include their work in the NASA CP.

Modeling and simulation-based engineering and science is rapidly becoming an essential scientific methodology for nearly all areas of engineering and many branches of science and for research, development, concept generation, product design and manufacturing, and consumer marketing. Continuing advances in computational science and networking technologies have made modeling and simulation-based engineering and science a “powerful and ubiquitous tool” for engineers and scientists and have made it possible to extend the range, depth, and applications of modeling and simulation vastly, especially when the investigated phenomena are not observable or measurements are impractical or too expensive. According to a National Science Foundation Blue Ribbon Panel, modeling and simulation-based engineering and science: 1) is an equal and indispensable partner, along with theory and experiment, in the quest for enhanced technological innovation; 2) holds great promise for the pervasive advancement of knowledge and understanding through discovery; 3) is indispensable to the Nation’s continued leadership in innovation and economic global competitiveness; and 4) is “key” to advances in a variety of fields – biomedicine, manufacturing, systems engineering, nanotechnology, health care, atmospheric and climate science, energy and environmental sciences, advanced materials, and product development. Modeling and simulation-based engineering and science is also essential to the success of NASA’s research and many of its missions and projects.

Thomas E. Pinelli, Ph.D.
NASA Langley Research Center

Leanna (Dee) Bullock
Alliant Techsystems, Inc., NASA Langley Research Center

Table of Contents

1.0	DEFENSE, HOMELAND SECURITY & FIRST RESPONDERS TRACK	1
2.0	HEALTH & MEDICINE TRACK	136
3.0	HUMAN DIMENSION TRACK	160
4.0	EDUCATION TRACK.....	305
5.0	SERIOUS GAMES & VIRTUAL WORLDS TRACK.....	338

1.0 DEFENSE, HOMELAND SECURITY & FIRST RESPONDERS TRACK

Track Chair – Steve Husak, National Training and Simulation Association



1.1	Emergent Capabilities Converging into M&S 2.0	2
1.2	JELC-LITE: Unconventional Instructional Design for Special Operations Training	9
1.3	Wargaming in Both Rectilinear and Hexagonal Spaces	17
1.4	Data Farming and the Exploration of Inter-agency, Inter-disciplinary, and International “What If?” Questions	27
1.5	High Resolution Visualization Applied to Future Heavy Airlift Concept Development and Evaluation	37
1.6	GPR Imaging for Deeply Buried Objects: A Comparative Study Based on FDTD Models and Field Experiments	45
1.7	Assessing Tsunami Vulnerabilities of Geographies with Shallow Water Equations	52
1.8	Command Center Training Tool (C ² T ²)	62
1.9	Disaster Response Modeling through Discrete-Event Simulation.....	68
1.10	Modeling a Civil Event Case Study for Consequence Management using the IMPRINT Forces module.....	74
1.11	Cognitive Systems Modeling and Analysis of Command & Control Systems	83
1.12	Colonel Blotto Games and Lancaster’s Equations: A Novel Military Modeling Combination.....	93
1.13	Mission Assurance Modeling and Simulation: A Cyber Security Roadmap.....	100
1.14	Framework for Assessing Cost and Technology: An Enterprise Strategy for Modeling and Simulation Based Analysis	109
1.15	A First Person Shooter/Real Time Strategy Hybrid: Lessons Learned	119
1.16	Framework for Architecture Trade Study using MBSE and Performance Simulation ..	127

1.1 Emergent Capabilities Converging into M&S 2.0

Emergent Capabilities Converging into M&S 2.0

Emilie Reitz
General Dynamics Information Technology
emilie.reitz.ctr@hr.js.mil

Jay Reist
Joint and Coalition Warfighting
Jay.reist@hr.js.mil

Abstract. The continued operational environment complexity faced by the Department of Defense, despite a restricted resource environment, is a mandate for greater adaptability and availability in joint training. To address these constraints, this paper proposes a model for the potential integration of adaptability training, virtual world capabilities and immersive training into the wider Joint Live Virtual and Constructive (JLVC) Federation, supported by human social cultural and behavior modeling, and measurement and assessment. By fusing those capabilities and modeling and simulation enhancements into the JLVC federation, it will create a force who is more apt to arrive at and implement correct decisions, and more able to appropriately seize initiative in the field. The model would allow for the testing and training of capabilities and TTPs that cannot be reasonably explored to their logical conclusions in a 'live' environment, as well as enhance training fidelity for all echelons and tasks.

1. INTRODUCTION

The future that our warfighters face will be complex in terms of both the geographical and human terrain [13]. This is a mandate for greater adaptability at lower tactical echelons to achieve operational objectives more efficiently. To prepare our warfighters to operate in this environment, greater access to embedded cultural and language training, engagement rehearsal capabilities, and realistic immersive environments are required.

This can be accomplished through the convergence of the fields of human and organizational adaptability; a persistent globally accessible virtual world; and, specifically, realistic virtual training environments that stress the warfighter on a physical and mental level. Combined, these integrated focus areas will make the training experience more holistically valid and representative of actual combat conditions. The resulting model of converging ideals will be supported by the implementation of independent assessment, including extensive measurement, best training practices from across the continuum, and cognitive learning theories.

1.2 Operational Problem

"In order to sustain operational and training superiority, the DoD must be able to effectively and efficiently prepare future

training audiences with limited fiscal, time, material, and personnel resources" [5, pg.3]. To accomplish this, the department must undertake a focused effort on creating and measuring adaptive human performance and learning methodologies for combat and combat support forces and apply it to training.

Over the last three decades, the academic and scientific communities have made great improvements in learning techniques and human performance – particularly advancements in learning under stress [1].

Through integrating state of the art human performance learning processes, stress inoculation techniques and procedures into the current modeling, simulation and training continuum, we can expect:

- Improved individual and team performance.
- Improved situational understanding, resulting in better decision making.
- Improved coping skills with the stress of combat.
- Capture, measurement, and assessment of training effectiveness.
- More confidence and ability to adapt correctly to a changing situation.
- A mechanism to inculcate best practices from experts to newer warfighters.

- A bias for positive action that translates into an overall increase in subordinate initiative.

1.3 Current and Future Operating Environment

To frame the operational problem, it is necessary to examine the expected operating environment. The August 2010 Commander International Security Assistance Force (COMISAF) Counterinsurgency Guidance succinctly frames the challenges of the future operating environment in its overall theme as “Learn and adapt” (pg. 1). It calls for our warfighters to exercise initiative and for leaders to empower their subordinates down to the lowest tactical echelon possible. The COMISAF Guidance also places importance on cultural understanding, thinking from the perspective of our international partners and our adversaries [2]. As brought up in the 2010 Quadrennial Defense Report, prevailing against our current adaptive adversaries requires a dynamic toolset of capabilities which not only focuses on military competence but also intelligence, diplomacy, and law enforcement activities [4].

Learning and adapting is not just a need in the current operating environment but a requirement to survive and prevail. In its current state, learning and adapting occurs as happenstance, not a systemic activity. The requirement for a dynamic, adaptive force will only increase in the future. The Joint Operating Environment (JOE) 2010 laid out constants and expected trends the Department of Defense will face for the next 15 years. The JOE’s constants weigh on the need for adaptability in our forces: there is the nature of war itself, filled with unexpected frictions; the nature of change, drastic as it can be; disruptions in the world and the battlefield, whether they are brought on by human error, natural disaster, or inflicted purposefully; and the flaws in maintaining grand strategies, despite the

challenge of disruptions and the nature of change [13].

One predicted change that will increase the challenge for the joint forces is the worldwide shift in urban density. The trends of continued resource scarcity, interdependent development, heightened technological evolution, and greater ease of pandemic spread interact with the shift of urban density, and will influence the ability of our joint forces to pursue missions. It is expected that by the 30s, five billion of the world’s total eight billion in population will reside in cities; two billion of those will be in the Middle East, Africa and Asia. The JOE 2010 goes on to state, “With so much of the world’s population crammed into dense urban areas and their immediate surroundings, future Joint Force commanders will be unable to evade operations in urban terrain” [13, pg. 57].

Historically, urban terrain requires more ground troops than any other type of terrain, while producing higher casualty rates for ground forces [7]. To maintain presence in a city, forces will have to act in concert with international partners and the host nation, which adds additional requirements for warfighters to maintain greater cultural understanding and refined communication skills [3]. The joint forces additionally will need a greater ability to understand second and third order effects. These skill sets need to be trained and honed in addition to standard warfighting tasks, and be provided from the same constrained pool of available time and resources.

The Capstone Concept for Joint Operations (2009) stresses that for the future, we must be prepared to “institute mechanisms to prepare general-purpose forces quickly for mission changes,” “develop innovative and adaptive leaders down to the lowest levels,” and “drive synergy to the lowest echelon at which it can be managed effectively” (pg. 21-28). This reinforces the distributed nature expected of future operations, where tactical troops will face strategic

compression, the compression of decision-making responsibilities down to the lowest level; and all echelons will be faced with the need to maintain a force that is organizationally flexible, able to respond quickly to chaotic changes in the operating environment. Training and technology can facilitate this shift towards quick forming teams of soldiers, multinationals and non-governmental agencies, as well as aid those units who are required to quickly transition from a combat mission to security or stability missions [3].

1.4 Mandate

The Strategic Plan for the Next Generation of Training for the Department of Defense states that it expects a revolution in training which “will be driven by the dual engines of powerful new training technologies and adaptive pedagogy that will fundamentally change how DoD Components train day-to-day at home station” (pg. 2) This will occur concurrently with the previously mentioned limited resources and encroachment on training grounds [5]. The solution is to augment live training with other capabilities. Immersive training permits for the realistic replication of any operating environment yet cannot exist in a vacuum of stand-alone developments. The ability of immersive training to cope with the inevitable constraints to resources is recognized by *The Strategic Plan for the Next Generation of Training*, which explicitly articulates the need for an immersive training environment that trains to a range of objectives with a focus that includes training in cognition, intuition, innovation and adaptive thinking. All of this needs to be accomplished while remaining technologically intuitive and agile.

2. EMERGENT CAPABILITIES

Emergent capabilities are technology and training enablers that can support improved warfighter training when linked with developments within the fields of human and organizational adaptability, virtual worlds, and immersive training, and areas yet to be developed.

2.1 Essential Elements

2.1.1 Virtual Worlds and Immersive Training

The immersive training and virtual worlds capabilities must overlap in terms of providing structured training and persistent terrain and human models. It would supply a common space, a persistent world built to simulate in large portion the real world, available 24/7 for home station and collective training for disaggregated entities. Virtual worlds additionally allow for the testing and training of capabilities and tactics, techniques and procedures that cannot be fully explored to their logical conclusions in a ‘live’ environment. Immersive training, whether it is performed with a man worn system, in military operations in an urban terrain facility, or with augmented reality technologies, should link into the same virtual world capability. This interconnectedness will assist in providing the necessary complex, sensory-realistic environment which stimulates cognition, intuition, innovation and adaptive thinking, while honing complex decision making skills [5]. Such an environment would be used to bring together disparate groups and expertise separated by logistically challenging conditions.

2.1.2 Human Factors

By placing emphasis on enhancing the cognitive performance of the human utilizing the system, technological capabilities can produce a greater increase in training that transfers accurately to the operational environment. There is a large body of literature on the potential of raising the expertise level of trainees through training focused on enhancing decision making [12], as well as building resiliency through training [1]. Transferring such effective experiential learning is reliant on more than psychological factors; however, because of the importance of model end state rather than visual enjoyment, effective training can be transferred through low-end technology means. The importance is the functional fidelity of experience which re-creates the

opportunity to mimic expert-level mental models [11].

The scenarios provided in a virtual environment must test trainees' assumptions not only about the actions they are immediately taking but also their deeper assumption sets and the second and third order effects of their actions [8].

2.1.3 World Class Trainers

Without a pool of skilled trainers to support the training in immersive and virtual worlds, the possibility of negative training is high. Branching scenarios need to be created or offered as a scaffold for leaders to build their own training scenarios. Leveraging best practices pulled from industry and academia, skilled trainers would be educated in the most effective and efficient methodologies to heighten transfer of experiential learning to their trainees.

In the future, after action review capabilities within an ad-hoc virtual world training session could be facilitated by a digital tutor system which is aware of prior trainee live and virtual performance through use of a persistent avatar card. This system would further support on the ground trainers by providing information on the progress attained and results of these ad-hoc sessions. It also has the potential to provide support for creating appropriate learning within scenarios based on units' experiences downrange.

2.1.4 Measurement and Assessment

Fully leveraging emergent capabilities includes measured performance and a better understanding of how we learn under stress as well as the heightened transfer of training that would be provided by an advanced training methodology. As we create, support and integrate technologies, solid metrics allow for true evaluation of the return on investment, as well as key areas for improvement as we develop the next system. Examples of measurements which will bring us towards the desired end state of institutionalizing adaptive virtual training

are: performance under stress, mission essential task performance, decision making performance, assessments of training practices, technologies, and trainer performance.

2.2 Building on Current Developments

The first steps towards integrating emergent capabilities into modeling and simulation of the future have already begun, as represented by the Joint Training and Exercise Network (JTEN) 2.0, and ongoing development expanding the JLVC Federation. These developments are bolstered by the US Army's Maneuver Center of Excellence's focus on bringing interconnectivity to the dismounted unit, while providing them a persistent, accurate virtual avatar for insertion into immersive and virtual training worlds [9]. The unification of systems that link live, virtual and constructive training with these more human-centric enablers will be key supporting capabilities in taking the first steps towards addressing the operational problem.

2.2.1 JLVC Federation

The JLVC Federation allows for the seamless integration of constructive and virtual entities into live training exercises, filling gaps in resources and spreading the training's reach to units who would otherwise be unable to participate. The functional fidelity provided by the JLVC Federation during training events is key, as many players were unable to tell that the air assets or nearby units with which they were coordinating were actually simulated by a computer located thousands of miles away [6]. This capability demonstrates the potential of uniting virtual worlds and immersive training.

2.2.2 JTEN

The backbone to this system is the JTEN, which is transitioning to an upgraded architecture with capacity to provide better speed, connectivity, and responsive interactions [10]. Most recently, the network was extended in a demonstration capacity

to provide support for tactical networks in the field that are operating out of the reach of the wire networks and support which accompanies most JTEN executions [6].

Supporting a persistent integrated immersive training capability available to all warfighters is crucial to achieve the vision of fully integrated training not only among the joint forces, but coalition as well [14].

3. CONCEPT FOR MODEL IMPLEMENTATION

These developments will not occur overnight, but rather progress incrementally. The Department of Defense can support these increments by focusing on creating an overall leap-ahead advancement in the technology. Technology in this case is defined in both materiel and non-materiel terms. This will be an iterative process to reach the end state, pushing the boundaries of available technologies and research with every iteration. Each stage would have a solid task statement and aggressive completion deadlines to speed delivery to the warfighters.

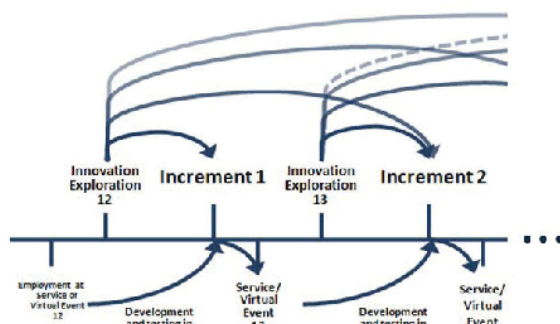


Figure 1. Developing Innovation.

While the terms increment and leap-ahead seem contradictory when placed side by side, they are actually mutually supportive in action (See Fig. 1). Each series of innovation exploration leads to the discovery of training and technologies that are then rolled into development increments at the appropriate time according to the Technology Readiness Level and cost. To accomplish some of the technologies which will bring us closer to the desired end state,

it will require longer periods of investment with industry, labs and academia. However, there are technologies available today or in the near term that need only to be integrated through incentivization to achieve the level of functionality needed. The incremental integration of these technologies into exercises and events will help industry and academia refine the utility of their demonstrated capabilities.

Key to accomplishing the model is identifying which parts of the growth process can be implemented in the short term, and which portions need to be worked and supported for inclusion into a later year's increment of development as part of the overall investment plan. Immature technologies and concepts are encouraged towards further development, while mature technologies and concepts are assessed and integrated quickly to maintain speedy delivery of capabilities to our forces. This approach will inform the technology associated with immersive training and virtual worlds, and spur development, research and integration of concepts.

3.1 Setting Conditions for Innovation

3.1.1 Innovation Exploration

Exploring innovation can be performed at existent major industry, entertainment and academic events, in collusion with partner organizations' exploratory events, or specifically tailored events hosted live or virtually. Products, capabilities and theories shared at these events are then analyzed for their current development level and applicability to future desired capability criteria.

In addition to directed exploratory events in real and virtual worlds, we would need to provide for targeted academic working groups and conferences, to allow for the further development of ideas and theories across disparate fields of study.

3.1.2 Technology Criteria

To qualify for movement from the technology exploration event to a development increment, technologies and training will need to demonstrate that they contribute substantially towards filling an identified gap.

3.1.3 Battle Labs

The creation of dedicated field and home-station based labs has aided previous fast-paced training and technology developments. Creation of a true battle lab for each service at a training site of their choosing would create further interoperability among the Services, industry, entertainment and academia while maintaining direct focus on the intended training audience.

This would allow for products from previous initiatives to be further refined and tested, and would allow immature innovations a location to refine their concepts. It would aid in initial scenario development, training and technology testing by leveraging actual service members.

4. CONCLUSIONS

In the complex environments we will continue to face, there is critical importance to being able to increase training capabilities offered to the warfighter, while improving their critical thinking and decision making. The Quadrennial Defense Report stated that one forward focus area is joint force holistic health. Citing programs such as the Comprehensive Soldier Fitness program, the Quadrennial Defense Report called out the importance of physical, psychological, spiritual and social health [4]. The ability to effectively make decisions (based on one's situational awareness, critical thinking, resilience, duress/stress, meta-cognition, sense making, and cue/pattern recognition) is part of joint force holistic psychological health. That skill set has a high degree of utility in the ongoing nature of strategic compression, which has increased the necessity for subordinates to appropriately seize initiative with an eye to second and third order effects. It requires

enhanced ability to receive and deliver precise information. To understand the implications of their actions, warfighters need to be able to leverage an adversarial point of view.

In the near future, adaptability, immersive training and virtual worlds will converge, touching not only enhancements in training and education, but also readiness across the force. Members of a high performance unit on patrol through a village will have the ability to articulate their understanding of changes in the local pattern of life, react to the dangerous situation before it becomes fatal, and act from a position of strength to manipulate their circumstances.

Because the team leader will wear a small, high resolution camera utilized to relay a live feed to higher headquarters, top trainers are able to capture that patrol and translate it into a scenario for the cross Department of Defense immersive training scenario pool. This allows for pre-deployment units to train in a high fidelity version of that original patrol, which accurately replicates the hair standing up on the back of the neck feeling experienced by the original patrol.

Since the event is loaded into the scenario pool, it is accessible for trainers, academics and the original team members. This access would allow for the creation of deeper lessons learned, manipulation of the scenario in a test bed environment, or allow team members to re-experience a physician-guided portion of the scenario as part of a post-traumatic stress disorder mitigation strategy.

By creatively leveraging such an intertwined capability, we create a force that is better able to cope with stress, more apt to arrive at and implement correct decisions, and more able to appropriately seize initiative in the field.

5. REFERENCES

1. Cannon-Bowers, J. A., and Salas, E. (eds). (1998). *Making Decisions Under*

- Stress: Implications for Individual and Team Training*. Washington, DC: American Psychological Association.
2. Commander International Security Forces/Commander US Forces Afghanistan. (2010). *COMISAF's Counterinsurgency Guidance*. Kabul, Afghanistan, HQ ISAF/USFOR-A.
 3. Department of Defense. (2009). *The Capstone Concept for Joint Operations, v3.0*. Washington, D.C.
 4. Department of Defense. (2010a). *Quadrennial Defense Review Report*. Washington, D.C.
 5. Department of Defense. (2010b). *The Strategic Plan for the Next Generation of Training for the Department of Defense*. Washington D.C.: Office of the Under Secretary of Defense (Personnel & Readiness).
 6. Kostoff, J.S. (2010). Live and Virtual Assets Train Jointly. *SIGNAL Magazine*, June 2010. Retrieved from http://www.afcea.org/signal/articles/templates/Signal_Article_Template.asp?articleid=2303&zoneid=296
 7. Holcomb, J. (2000). Medical Support. In Glenn, R. (Ed.), *Capital Preservation: Preparing for Urban Operations in the Twenty-First Century*. Proceedings of the RAND Arroyo-TRADOC-MCWL-OSD Urban Operations Conference, March 22-23, 2000.
 8. Greitzer, F. L. & Andrews, D.H. (2010). Training Strategies to Mitigate Expectancy-Induced Response Bias in Combat identification: A Research Agenda. In D. H. Andrews, R. P. Herz, & M. B. Wolf (Eds.), *Human Factors in Combat identification* (pp. 173-189). Burlington, VT: Ashgate Publishing Company.
 9. Little, V. (2011). "MCoE boosts 'squad as a strategic formation'." *The Bayonet*, June 15, 2011.
 10. McCormick, B. (2011). DISA and USJFCOM Support warfighter Training. *The Grid*, Spring 2011. Retrieved from <http://www.disa.mil/news/grid/spring2011/usjfc.com.html>
 11. Means, B., Salas, E., Crandall, B. & Jacobs, T.O. (1993). Training decision makers for the real world. In G. Klein, J. Orasanu, R. Calderwood & C.E. Zsombok (Eds.), *Decision-making in action: Models and methods* (pp. 306-326). Norwood, NJ: Ablex.
 12. Spiker, V. A., & Williams, G. (2010). Summary of Interviews with Combat Hunter Graduates from the 1/5. Orlando, FL: Naval Air Warfare Center Training Systems Division.
 13. United States Joint Forces Command. (2010). *The Joint Operating Environment (JOE) 2010*. Norfolk, VA: United States Joint Forces Command.
 14. Walrond, J.T. (2009). Joint Training with International Partners: A Key to Achieving the Capstone Concept for Joint Operations. *Military Training Technology*, Vol 14:3. Retrieved from <http://www.military-training-technology.com/mt2-home/109-mt2-2009-volume-14-issue-3/953-usjfc.com-update.html>

6. ACKNOWLEDGMENT

This work was supported in part by the US Joint Forces Command (**Contract # N65236-09-D-3809**). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the USJFCOM or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

1.2 JELC-LITE: Unconventional Instructional Design for Special Operations Training

JELC-LITE: Unconventional Instructional Design for Special Operations Training

Gerald “Jay” Gendron
gerald.gendron@gmail.com

Mark Friedman, CKMP
Concurrent Technologies Corporation
friedmam@ctc.com

Abstract. Current special operations staff training is based on the Joint Event Life Cycle (JELC). It addresses operational level tasks in multi-week, live military exercises which are planned over a 12 to 18 month timeframe. As the military experiences changing global mission sets, shorter training events using distributed technologies will increasingly be needed to augment traditional training. JELC-Lite is a new approach for providing relevant training between large scale exercises. This new streamlined, responsive training model uses distributed and virtualized training technologies to establish simulated scenarios. It keeps proficiency levels closer to optimal levels – thereby reducing the performance degradation inherent in periodic training. It can be delivered to military as well as under-reached interagency groups to facilitate agile, repetitive training events. JELC-Lite is described by four phases paralleling the JELC, differing mostly in scope and scale. It has been successfully used with a Theater Special Operations Command and fits well within the current environment of reduced personnel and financial resources.

1.0 INTRODUCTION

Unconventional

The term *unconventional* not only describes a classic mission set of special operations forces – it also describes the manner in which special operations trainers are approaching training development in this time of constrained resources and changing missions. Special operations trainers have begun to use rapidly developed, distance training ways and means to surmount the resource challenges and achieve their mission goals.

As special operations and conventional forces continue to realize the changing global environment – and the resultant revision in their global mission sets – enabling functions like mission support, acquisition, and training should likewise change to keep in step with the needs of the men and women in uniform. From the perspective of joint military training, the revised mission sets will drive a requirement for increases of relatively short training events which augment the traditional joint training currently designed, planned, and

delivered by the Joint Staff and military Services.

Personnel constraints are perhaps the most challenging hurdle in training development. Joint military training is inherently a humanistic endeavor. It relies on a very particular and special subset of the profession of arms – the ability to discern training needs of operational, uniformed members and then translate those needs into specific, achievable, and timely training events. These events often include several hundred live players, not to mention the virtual players linked from distributed locations and synthetic forces generated through constructive simulations. People and their specialized training skills continue to serve as the most important element of the joint military training enterprise spread across the globe – the art and science of war. As a support function, they have taken resource cuts in order to bolster operational forces engaged in our nation’s wars. Yet the operational forces need even more training than in past years to deal with force surges, increased uncertainty, and changing mission sets.

The complexity of the current and foreseeable defense and security operating environments call for adaptive individuals, leaders, and units at all levels. While acknowledging that these characteristics already exist in the armed forces (to some degree), it is clear that the severity of the challenges imply an urgent need to significantly reinforce and extend them. We have entered a time when the uniformed services must be able to train more efficiently and with greater agility than ever before.

2.0 MILITARY TRAINING DEVELOPMENT ENVIRONMENTS

2.1 Key Terms and Concepts

Each year, the Department of Defense embarks upon a very challenging and broad-reaching joint training mission. Fortunately for our men and women in uniform, the Department continues to perform well in this mission. It has continued to enhance its operational-level training events through new methods and technologies. This paper focuses on training at the *operational level of war* which encompasses the forces supporting the *combatant commanders* around the globe.

From the *Department of Defense Dictionary of Military and Associated Terms* [1], *operational level of war* means “the level of war at which campaigns and major operations are planned, conducted, and sustained to achieve strategic objectives within theaters or other operational areas” ([1], p. 271). From the same source, the term *combatant commander* is defined as the “commander established and so designated by the President, through the Secretary of Defense and with the advice and assistance of the Chairman of the Joint Chiefs of Staff. Combatant commands typically have geographic or functional responsibilities” ([1], p. 60).

Most often, operational-level training results in the military Services working in concert with the Joint Staff to establish a training environment satisfying the training requirements. These are highly complex events, but the Joint Staff has promulgated policy and instructions in order to maintain effectiveness of the resultant training while striving for efficiencies. A key process in joint training is the Joint Event Life Cycle (JELC). The *Joint Training Manual for the Armed Forces of the United States* [2] defines JELC as:

A flexible sequential set of processes that can be modified to apply to various levels of joint event intensity. Although, nominally, a 12-18 month undertaking for major collective exercises, the JELC can be utilized for any training event and the planning team needed to accomplish the JELC is directly related to the scale and complexity of the training event itself ([2], p. E-6).

This is a foundational concept of the joint military training system and serves as the primary construct for this paper.

2.2 Joint Event Life Cycle in Military Training Events

Special operations training is not unlike conventional training from a process perspective. In fact, special operations trainers utilize the same JELC principles and timelines. At the operational level, the JELC specifically addresses operational-level tasks and results in exercises which typically last one week, though some are two weeks in duration. These exercises (i.e., training events) prepare thousands of members of the Armed Forces and the Department of Defense for the challenges of fulfilling its national security strategy around the globe. These exercises typically require a 12 to 18 month JELC to design, staff, and build the final event. As a result, most combatant commands have one, or possibly

two, full-scale training events each year. This flows from the personnel and schedule constraints of a Joint Staff required to support all the combatant commands.

2.3 Recent Occurrences Underpinning Change

Continued themes surrounding national security and the Department of Defense include uncertainty and change. In August 2010, Defense Secretary Robert Gates announced his intent for an efficiencies initiative to review the fiscal and personnel levels of the Department. The stated goal was to shift resources – approximately \$100 billion – to units and commands more directly supporting the war fighting missions. The shift would come from reductions in command staffs which grew, according to Secretary Gates, during the war years since 2001.

As these efficiency measures have been executed, the Obama Administration has noted the need to go further in an effort to reduce national debt. Efficiencies once focused on a shift of resources has transformed into a call for a reduction in Defense Department budgets by approximately \$400 billion through 2023 [3]. Consider the situation highlighted in the introductory comments of this paper:

- changing global missions demanding training support;
- constrained training personnel face increased demand for support; and
- training is an inherently humanistic endeavor.

The Joint Staff has foreseen these future challenges and has supported the development of new training technologies and methods to surmount them. These new technologies have been in immersive learning environments. They have been developed and are hosted on Joint

Knowledge Online for anyone in the Department of Defense to use for their training needs. Immersive learning environments are a step beyond computer based training. Rather than focusing on an individual, the technologies allow small teams to train on staff processes similar to those found at the operational level of war.

3.0 ADAPTING PEOPLE AND PROCESSES

The human aspects of quality training preclude a total solution derived solely from technologies alone. A truly effective training solution requires not only the technologies, but also the right people and processes.

3.1 New Approach: More in Less Time

The JELC-Lite model proposed offers two major benefits which make it ideal to supplement the traditional JELC process: reduced development time for training events, and reduced regression of staff skills.

Traditional joint training events have durations lasting approximately one week. In contrast, collaborative staff planning events have an average duration of four hours when using immersive learning environments. These shorter events allow operational commanders and training officers to increase the frequency of events. Typical outcomes of these four-hour events include: rehearsing existing staff knowledge, expanding the knowledge base from veteran staff to new team members, and building (nurturing) relationships among collaborative teams.

JELC-Lite also reduces the amount of time to develop training events on a per training-hour basis. As noted, the JELC operates over a 12 to 18 month timeframe to design and build a one week training event. This typical situation results in a development-

delivery ratio of 52:1 (52 weeks:40 hours). Experience has shown that JELC-Lite events can be designed and created in approximately 2 weeks. This results in a development-delivery ratio of 20:1 (2 weeks:4 hours). The JELC-Lite performance ratio is 62 percent more efficient as compared to the JELC (i.e. requires 62% less labor to create an hour of team training). This makes JELC-Lite training processes nearly three times quicker than traditional ones.

Joint observer-trainers note staff performance is heightened by its participation in a one-week event on an annual basis; however, the staff performance level steadily declines over the course of the intervening year until the next event. The need for long duration, comprehensive training events using JELC processes will persist, but JELC-Lite provides the opportunity to augment these yearly events with short duration training opportunities. Fortunately, the guidance from the authoritative source on joint training not only allows adaptation of the JELC but also encourages it to “fit the scale and intensity of the training event itself” ([2], p. E-6).

Adapting JELC for JELC-Lite will allow the execution of relevant training to learners, both before and between the large-scale exercises, in order to maintain higher levels of proficiency with greater consistency – avoiding the bathtub effect of decreased performance that often occurs between large-scale events. Not only are staffs more capable, but training events are more beneficial when the regression of staff skills is lessened.

JELC-Lite results in “a lot of a little” – an idea based on the book *The Long Tail: Why the Future of Business Is Selling Less of More* by Chris Anderson. Anderson [4] describes how the Internet revolution has been shaped by a specific nuance that

technology has afforded the learning industry. Learners can now focus on a very narrow and specific topic, instead of having to rely on large, broad, and generalized events. Not disparaging these larger, aggregated training events – but the opportunity to narrowly-focus and train on specific mission essential tasks provides the trainer flexibility to create and deliver many times more training events which are much more easily adaptable to the ever-changing set of requirements in today’s uncertain environment.

3.2 Four Phases of Proposed JELC-Lite

Traditional JELC has five defined stages [2]:

- Design
- Planning
- Preparation
- Execution
- Evaluation, Analysis, and Reporting

JELC-Lite has four phases, which closely mirror the standard Joint Event Life Cycle:

- Concept Development
- Plan-Build
- Pre-Execution Preparation
- Execution

The primary differences between JELC and JELC-Lite lay in the scope and scale, with the Lite version being more responsive and agile. It has been expressly designed for use with immersive learning environments capable of rapid training scenario development. The four phases are described below. A guide detailing the JELC-Lite process is being written by the authors to provide training developers and training audiences guidance on employing JELC-Lite in their events.

3.2.1 Concept Development

Concept Development has three main purposes: establish expectations, develop initial action items, and finalize the requests for support from customer to training development team. Most important is setting proper expectations – especially when developing training under JELC-Lite. There is a longstanding understanding of traditional JELC processes. Users must appreciate that a rapid development effort will require a different approach to building the event under JELC-Lite.

The focus needs to be on the training audience and their ability to play in the event rather than the technologies used. An example of the checklist used during this phase: the developers facilitate a session wherein customers determine event objectives, identify a geographical location for the scenario, determine the number of hours for event execution, and scope the number of scenario injects the players will have to manage.

3.2.2 Plan-Build

Where JELC typically has four planning conferences, the JELC-Lite Plan-Build Phase is performed only once, focusing on a small target set of objectives. The primary objectives in this phase of JELC-Lite are for the customer to decide on the player roles (how many? how broad?); to define the activities within the event (one planning activity? planning with a fast forward to deployment?); and to storyboard the activities and roles against the desired training injects and synchronize timing. The Plan-Build Phase co-exists with a series of product reviews at 30 percent (review inject products), 70 percent (review overall event flow), and 90 percent (review details and synchronization). Recall that this phase may last only a week, so these reviews help drive the process through the longest portion of development. By leveraging

toolset capabilities like those in the Small Group Scenario Trainer (SGST), visual diagrams can be created, shared, edited, and finalized over web services [5]. Aristotle said it best: *“The soul does not think without a picture.”*

3.2.3 Pre-Execution Preparation

Having drafted and constructed the training event, this phase is designed to rehearse the event with the customers to ensure expectations for players, controllers, and role players are satisfied. Pre-Execution Preparation can be accomplished in the JELC-Lite model via simple email and teleconferences. This phase is very important in the distributed, lightweight JELC-Lite model, since face-to-face coordination may not be possible during rapid development. A side benefit is that it reduces loads on personnel and fiscal resources as compared to traditional preparation under JELC.

3.2.4 Execution

Lastly, the Execution Phase includes scenario delivery, a *hotwash* with the training audience, and an internal lessons learned session among the developers. These are very similar to the activities in a traditional JELC process.

4.0 A PROMISING VIEW OF THE FUTURE

JELC-Lite processes have been successfully used at a few combatant commands from 2010 to the present. Three overarching benefits have been observed among training developers and audiences employing JELC-Lite.

4.1 Expand Training Reach and Responsiveness

JELC-Lite supports the philosophy “a lot of a little”. This means not only a greater

number of events, but with the correlating effect resulting from more frequent events. Customers can avoid having to wait for the next large-scale exercise to rehearse or train techniques, tactics, and procedures. In fact, now operational commanders can have their staffs go through those experiences as a build up for the large-scale event.

Training can become not only more frequent but also more penetrating in developing our human capital. There exists a subtle (yet significant) side benefit to this increased reach – increased responsiveness.

Developers and analysts involved with training recognize that the 18 month time spans needed to develop large-scale events are more susceptible to obsolescence due to changes in global environments that alter the mission focus. More frequent event development and training reduces risk in an uncertain operational environment.

4.2 Agile Solution

The responsiveness inherent in reduced JELC-Lite development cycles fosters agility for the training audience. In today's constrained resource environment, there should be a shift in thinking that large and situationally-specific scenarios will be reused in the future to gain economies of scale. Rather, training developers can espouse the benefits of situationally-specific scenarios that are smaller in scope and easier to develop in order to remain agile to training needs – and ultimately cost less.

"The need to learn quickly and adapt in a dynamic environment is seminal for both military and civilian organizations" ([6], p. 165).

The added benefit of this agility is that the trainers are able, and encouraged, to update, revise, and improve the training content between each cycle of delivery. Since there is no end-state to the idea of "better" and "continuous improvement" – JELC-Lite provides the opportunity to

sustain the idea flow. In other words, the shorter cycle time aids the addition of innovation into the training. Without the shortened cycle time for training development, there would simply be less opportunity to achieve this level of innovation.

4.3 Evolutionary Approach

Finally, this new streamlined, responsive, and agile training model called JELC-Lite affords trainers the opportunity to reach under-reached training audiences and allows for agile, repetitive training events to take place where needed. This mimics a popular paradigm found on the Internet in the commercial sector – that being Apple's iTunes delivery capability. Consider the iTunes inventory: it is a huge assortment of media which is instantly available to customers. Meanwhile, the iTunes process is highly repeatable, has very low transactional cost, and offers very efficient management overhead. Thus, the evolutionary adaption of JELC-Lite to augment the JELC could be whimsically called iJELC.

Who are the under-reached and under-served training audiences? There are four levels of penetration when expanding reach to users:

- Active Duty Military Forces;
- Reserve and National Guard Forces;
- Interagency Partners; and
- Multinational Partners.

Among active duty users, consider newly arriving members of a joint, operational staff. A study [7] commissioned in 2008 by the Joint Staff J7's Joint Training and Exercise Division found officers serving on operational-level joint staffs were not performing as well as senior leaders had expected. Reasons for this cited by the study include the facts that these officers were on their first joint tour, their first staff

tour, and their first time working with a diverse workforce such as other military Services and the interagency. Experience with operational joint staffs points to a lack of process-oriented training on staff techniques and skills. Newest members of a combatant command must seek on-the-job training in preparation for the large-scale, annual training exercise. Newly assigned members of combatant commands and the Joint Staff represent an under-served audience during their first four to eight months of time on the staff.

Secondly, consider the Reserve and National Guard forces who volunteer to support active duty units. Most frequently in the special operations community, these groups of warfighters augment the active duty components by deploying on multi-month Joint Special Operations Task Force tours. While they are often able to take long blocks of time from their civilian jobs to support the military mission, they can have difficulty getting additional, smaller blocks of time away from work prior to the deployment to support important training workup events as preparation for the deployment. This results in a less than fully prepared group of staff members who could benefit from collaborative staff training using distance training tools and techniques prior to deployment.

Beyond the uniformed members, there has been a notable increase in cooperation, engagement, and mission execution among the entire interagency including Department of Defense, Department of State, law enforcement agencies, and more than a dozen other governmental agencies. One persistent problem cited by non-defense members of the interagency is the inability to train with uniformed staff members. There are just so much fewer personnel in non-defense agencies and this leads to difficulties to engage in training with the military – especially when travel is required. The Department of Defense has many more

members than the rest of the interagency, but uniformed military members often fail to realize this disparity in personnel resourcing. Again, these other members of the interagency could benefit from collaborative staff training using distance training tools and techniques. It would increase interaction among key mission partners.

Finally, the military supports the Department of State diplomatic mission by participating in Security Assistance activities like training and educating our multinational partners. A key problem is a lack of persistence in contact due to costs and distance challenges of working with multinational partners. Additionally, not all our partners are supported at similar levels. This is to be expected. However, when it comes time for execution it is not desirable to expect less of a less-served partner. Fiscal and personnel constraints make it difficult, if not impossible, for the Department of Defense trainers to travel and support these multinational partners frequently enough to maintain the performance levels achieved after the live security assistance training. Distance training and a JELC-Lite process may bring recurring opportunities to smaller, lesser-served – but nonetheless important – partners.

5.0 CONCLUSION

Today's operational-level training events are designed, constructed, and executed under a well-established process called the Joint Event Life Cycle (JELC). Special operations training events use the JELC as well. The JELC typically requires 12 to 18 months to plan a one week joint training event. However, a changing landscape in national security, and tightening constraints on critical resources, such as people and money, requires the Department of Defense to reconsider guidance in CJSCM 3500.03C stating it is appropriate to adapt the JELC to

the scale and complexity of the desired event.

As the military realizes changing global mission sets, shorter training events using distributed technologies will increasingly augment traditional training. JELC-Lite is a new approach to provide relevant training between large scale exercises. This paper articulates the four phases the authors have defined for JELC-Lite and the benefits JELC-Lite has for operational commands. Overall, a greater number of short training events points to a future where higher performance levels are maintained between large-scale training events. Maintaining more optimal proficiency levels reduces the performance degradation inherent in periodic training. This new streamlined, responsive training model uses distributed and virtualized training technologies to establish simulated scenarios. It can be delivered to military as well as under-reached interagency groups to facilitate agile, repetitive training events.

6.0 REFERENCES

- [1] Department of Defense. (2011, May 15). *Department of defense dictionary of military and associated terms* (JP 1-02). Washington, DC.
- [2] Department of Defense. (2011, January 15). *Joint training manual for the armed forces of the united states* (CJCSM 3500.03C). Washington, DC.
- [3] Winter, D.C. (2011, May 31). *Defense planning for the twenty-first century: Lessons from the twentieth* (Report of the Heritage Foundation). Retrieved from <http://www.heritage.org/research/reports/2011/05/defense-planning-for-the-twenty-first-century-lessons-from-the-twentieth>.
- [4] Anderson, C. (2006, July 11). *The long tail: Why the future of business is selling less of more*. New York, NY: Hyperion.
- [5] Joint Knowledge Development & Distribution Capability. (2009, August 31). *Small group scenario trainer (SGST): Users guide*. Suffolk, VA.
- [6] Mains, S., & Ariely, G. A. (2011). Learning while fighting: Operational knowledge management that makes a difference. *PRISM*, 2 (3).165-176.
- [7] United States Special Operations Command. (2010, October). *US special operations command joint staff officer site visit*. Tampa, FL.

1.3 Wargaming in Both Rectilinear and Hexagonal Spaces

WARGAMING IN BOTH RECTILINEAR AND HEXAGONAL SPACES

Alex Hoover
Joint Coalition Warfighting
alex.hoover@hr.js.mil

Abstract. There are two main approaches to managing wargame entity interactions (movement, line of sight, area of effect, etc.): freespace and gridded. In the freespace approach, the units exist as entities in a continuous volume of (usually) Cartesian 3D space. They move in any direction (based on interaction with "terrain" that occupies the same space) and interact with each other based on references and displacements from their position in that space. In the gridded approach, space is broken up into (usually regular) shaped pieces. Units are considered to occupy the entire volume of one of these pieces; movement, line of sight, and other interactions are based on the relationships among the spaces rather than the absolute positions of the units themselves. Both approaches have advantages and drawbacks. The general issue that this discussion has addressed is that there is no "perfect" approach to implementing a wargaming battlespace. Each of them (and this extends to others not discussed) has different sets of advantages and disadvantages. Nothing will change that basic nature of the various approaches, nor would it be desirable to do so. Along with the advantages, the challenges define the feel of the game and focus the thinking of the players on certain aspects and away from others. The proposed approach to combining square and hexagonal approaches, which we will call the rhombus interface, leverages rhombuses constructed from equilateral triangles into which the hexagon can be decomposed to bridge the gap between the approaches, maintain relative consistency between the two as much as possible, and provide most of the feel of the hexagonal approach.

1.0 INTRODUCTION

There are two main approaches to managing wargame units interactions with their environments (movement, line of sight, area of effect, etc.): freespace and gridded. In the freespace approach, the units exist as entities in a continuous volume of (usually) Cartesian 3D space. They move in any direction (based on interaction with "terrain" that occupies the same space) and interact with each other based on references and displacements from their position in that space. In the gridded approach, space is broken up into (usually regular) shaped pieces. Units are considered to occupy the entire volume of one of these pieces; movement, line of sight, and other interactions are based on the relationships among the spaces rather than the absolute positions of the units themselves. Both approaches have advantages and drawbacks.

The freespace approach seems to be more "natural" or "realistic" from an intuitive sense. Units can occupy a number of positions and orientations that is only limited by the degree of effort you are willing to put into measurement. It is easy to make analogies between the physical world and the wargame battlespace.

The challenges with freespace are mainly in complexity and level of effort in implementation. Calculating relationships among points and volumes in freespace can be difficult, especially when dealing with paths through that space that are non-linear (they "go around" things in the space). There are a lot of physical techniques that can alleviate these challenges, mostly due to the close correspondence to the physical world. These techniques, however, usually increase the level of effort challenges inherent in the freespace approach. The archetypical level of effort challenge puts the player in the position that since they have so many degrees of freedom in action, that to gain best advantage in those actions, they have to put large amounts of effort into examining each action in a high level of detail. For example, an individual combatant who is walking a "S" path in freespace can have a significantly different advantage of position by making minute changes to the curvature and inflection points in the path.

The gridded approach is generally seen to overcome the challenges of freespace. By dividing the space up into reasonably sized discrete *cells*, movement, line of fire, cover, and other interactions can be implemented

with relatively simple systems based on whole number counting. Additionally, since there are no different positions within the cells, there are fewer degrees of freedom in actions, and thus the required or desired level of effort (“Waitaminute! If I moved my guy a half a millimeter to the right, would that let me ...”) in considering options is lessened. This decreased tactical focus frees up players’ cognitive resources for other, generally considered “deeper,” considerations such as “look-ahead,” strategy, and consideration of the meta-game.

These advantages do come at a cost. The reduction in degrees of freedom equates to a reduction in verisimilitude. For example, units can occupy a certain cell or another adjacent cell, but cannot occupy a space “in between” the cells that might logically equate to a tactical advantage over either, or be the outcome of a more naturalistic decision process such as the desire to stop movement directly “behind” a pillar that acts as an obstruction to fire rather than “to the right” or “to the left” of it. It is frequently the case that in three dimensional gridded systems one of the dimensions (usually “altitude”) is treated differently than the other two (forming the “ground”). This usually introduces complexity (in the form of additional rules) and again detracts from the verisimilitude.

There are other, more basic problems with the “feel” of gridded systems from the constraints imposed by the cell structure. In square grids moving four units *north* and then three units *east* results in taking seven unit steps, but the measured distance between the centers of the starting and ending cells is only five units. Breaking this movement up into a sequence of smaller movements does not alleviate the problem.

Hexagonal grids can mitigate this problem with square grids, as the correspondence between the minimum number of steps between cells and the measured distance is much closer. Additionally, hexagonal grids

provide six degrees of freedom in movement vice the four of square grids at no increase in rule complexity. Even though there are more options, they tend to seem less “natural,” possibly because people are more familiar with rectilinear grids and have more patterns of mind based on “forward, backward, right, left” as opposed to “every sixty degrees.” The biology of bilateral symmetry may even have an effect on this, which would be a difficult part of the mindset to overcome.

An additional issue with hexagonal grids is that they are not closed under composition, that is, you cannot take a regular hexagon and break it up into any whole number of smaller hexagons that completely cover the same shape. Square grids, however, are easily decomposed into smaller square grids, or aggregated into larger ones.

The general issue that this discussion has addressed is that there is no “perfect” approach to implementing a wargaming battlespace. Each of them (and this extends to others not discussed) has different sets of advantages and disadvantages. Nothing will change that basic nature of the various approaches, nor would it be desirable to do so. Along with the advantages, the challenges define the feel of the game and focus the thinking of the players on certain aspects and away from others.

2.0 A SOLUTION

Directing players’ thinking and behavioral patterns is a very important factor in the enjoyment, catharsis (immersion), and experiential learning value of wargames. While the approach to implementing the battlespace is not the only driver of the players’ experiences, it is a significant one since it affects most decisions and other thinking about the game. The goal of the proposed schema is to provide an approach to integrating square and hexagonal grids to make a hybrid battlespace that will allow the game to provide players access to the feel

of either system, in portions where appropriate, with a minimum of difficulties in transition between the two.

So what are the general feels of the square and hexagonal systems and what is a reasonable use case where you would want both combined?

The square gridded environment provides a *best fit* for man made environments, especially the interiors of structures and built up urban areas. The rectilinear structure of streets and hallways can usually be lined up with the grid itself, creating a natural harmony between the referent and the representation. In general, the more densely built up a man made environment is, the less opportunity there is for “diagonal” movement, so the less important the challenges with moving large distances at angles not aligned with the grid are. This system still has significant challenges representing large open areas and structures with multiple grids, offset angularly from each other.

As a sort of compliment, the hexagonal grid is most popular with outdoor environments. It provides the additional degrees of freedom of motion expected in open areas and has much less separation between steps and measured distances over larger portions of the environment. Likewise, the hexagonal system has significant challenges representing rectilinear shapes. In general, they have to be constrained to unnatural, offset positions and the hexes that contain the edges of the rectilinear shape usually have some *loss* – large portions of area that are inaccessible to the units in the game, but not actually occupied by the represented terrain itself.

One of the most common methods to bring these two approaches together is to not bring them together. Or, more specifically, to conduct different parts of an overall scenario on different maps that have entirely one or the other type of grid, as appropriate. The second most popular

method is to severely limit one or the other type of terrain on the map and allow it to be represented with the approach appropriate for the dominant part of the terrain. The entry point to the city might have a small strip of outdoor terrain at one edge of the map or the number houses and outbuildings in a rural area might be minimized. Both of these methods work well enough, but are certainly compromises to the idea of a hybrid square and hexagonal grid. The easy way, just mashing the two grids together and accepting or artificially putting obstructions in the inconvenient places where the two are not compatible is also somewhat unsatisfying.

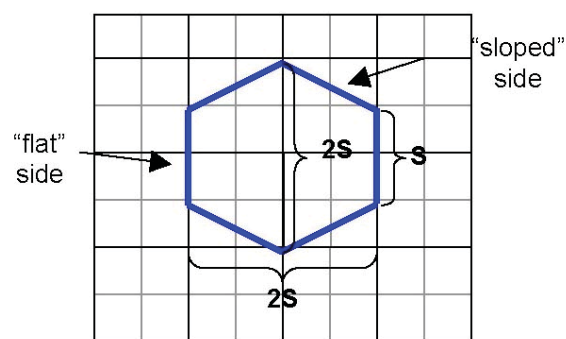


Figure 1. The “Stretched” Hexagon

The proposed approach to combining square and hexagonal approaches, which we will call the rhombus interface, leverages rhombuses constructed from equilateral triangles into which the hexagon can be decomposed to bridge the gap between the approaches, maintain relative consistency between the two as much as possible, and provide most of the feel of the hexagonal approach.

Figure 1 illustrates the foundational geometry of the approach. The hexagonal grid to be integrated will be made of slightly “stretched” hexagons (not quite regular hexagons), adjusted to mesh with a square grid of side $s/2$. One pair of opposite sides for these hexagons will be of length s and parallel to one pair of sides of the square grid and spaced $2s$ apart from each other, with end points aligned. We will call these the *flat* sides. The point-to-point diagonal of

the hexagon between the flat sides will also be of length $2s$, and will be in the middle of the flat sides, centered on their length. The sides of the hexagon that connect the flat sides to this diagonal will be called the *sloped sides*.

Note that by arranging one of the flat sides with one side of the square of the base grid, the points of the point-to-point diagonal that defines the sloped sides align with the half grid. Also note that the area of this hexagon is exactly $3s^2$, or the area of three of the full squares of the base grid.

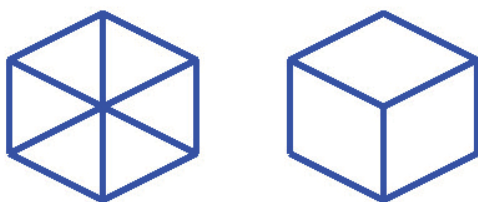


Figure 2. Triangles and Rhombi

The vertices of the hexagon align with a whole number ratio of the base grid, with no need to approximate irrationals. The half-unit is a standard subdivision of both English and Metric measures, one which is fairly easy to physically approximate without measuring. This makes calculation, vector construction, graphical representation, hand drawing of the grid, and calculation of an individual hexagon from a reference point without calculation of all the intervening hexagons all relatively easy and computationally simple. These properties will propagate across the grid as the hexagons are repeated.

Next, we will break this hex down into six triangles, then aggregate them to make the rhombi that will be the base cell for the organic part of the grid. As illustrated in Figure 2, the hexagon can be divided into congruent triangles by connecting each vertex to the center of the figure. Adjacent pairs of triangles are aggregated into rhombi by eliminating their common side. The resulting four sides make a rhombus. There

are two different ways you can do this for each hexagon. The one in Figure 2 will be called a *top set*, the opposite will be called a *bottom set*.

Since each of the three rhombi are congruent (each is composed from a pair of congruent triangles), it is easy to see that each has an area one-third of the original hexagon, or s^2 , the same area as the squares in the base grid. Also, the lengths of the pairs of sides are s for the flat sides (parallel to the flat hexagon sides) and a little less than $1.2s$ ($\sqrt{5}/2 s$, to be precise) for the sloped sides (again, parallel to the sloped sides of the hexagon).

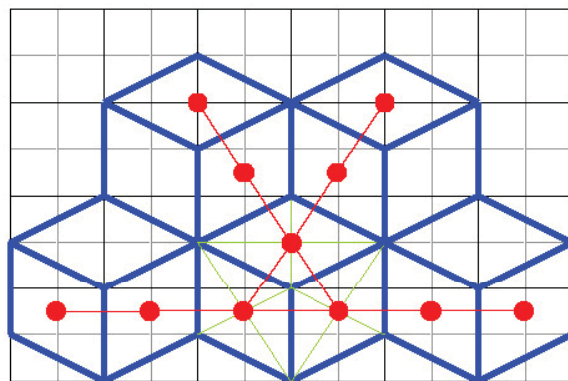


Figure 3. Cell to Cell Transitions

The rhomboid cell, then is decently close in basic size and shape parameters to the square grid cell and fairly simple to construct from that underlying grid. These are nice descriptive parameters, but the real key performance parameters are how the grid supports interactions among units in different cells.

3.0 DISCUSSION

3.1 Resolution and Distance

Qualitatively, we can see that the rhomboid grid gives us similar degrees of freedom for cell to cell transition as the square grid – four choices by passing into an adjacent rhombus that shares a flat (in some cases) or a sloped (in all cases) side. In the

aggregate, however, the cell to cell transition also replicates the hexagonal system, giving us six “straight line” directions overall, as illustrated in Figure 3.

Quantitatively, the cell to cell distance transitioning through a flat side is exactly s . The cell to cell distance transitioning through a sloped side is $\sqrt{13/16}s$, or a little over $.9s$, for just less than a 10% loss for steps vice geometric distance. When you

consider the effect on multiple transitions that change direction, the variances get larger, but also become dependent on path selection to get from one cell to another. A good metric to evaluate the overall distortion over larger distances is to calculate the relative area of a geometric circle of radius x and compare it to the sum of the areas of the unique cells that can be reached by x cell to cell transitions.

Table 1. Comparison of Areas

steps	circle area	square grid area		hexagonal grid area		rhomboid grid area	
1	3.14	5	159.15%	6.06	192.97%	5	159.15%
2	12.57	13	103.45%	16.45	130.94%	13	103.45%
3	28.27	25	88.42%	32.04	113.33%	27	95.49%
4	50.27	41	81.57%	52.83	105.10%	43	85.55%
5	78.54	61	77.67%	78.81	100.34%	65	82.76%
6	113.10	85	75.16%	109.99	97.25%	93	82.23%
7	153.94	113	73.41%	146.36	95.08%	123	79.90%
8	201.06	145	72.12%	187.93	93.47%	161	80.07%
9	254.47	181	71.13%	234.69	92.23%	199	78.20%
10	314.16	221	70.35%	286.65	91.24%	247	78.62%
11	380.13	265	69.71%	343.81	90.45%	293	77.08%
12	452.39	313	69.19%	406.17	89.78%	351	77.59%
13	530.93	365	68.75%	473.72	89.22%	405	76.28%
14	615.75	421	68.37%	546.46	88.75%	473	76.82%
15	706.86	481	68.05%	624.40	88.34%	535	75.69%
16	804.25	545	67.77%	707.54	87.98%	613	76.22%
17	907.92	613	67.52%	795.88	87.66%	683	75.23%
18	1017.88	685	67.30%	889.41	87.38%	771	75.75%
19	1134.11	761	67.10%	988.13	87.13%	849	74.86%
20	1256.64	841	66.92%	1092.06	86.90%	947	75.36%

The above table shows the area of a circle in free space as compared to the area of the “same size” circle as it can best be approximated in grid space, hexagonal space, and rhomboid space. As you increase the number of steps (which can represent either the distance travelled or the chosen grid resolution), the implementations approach a constant ratio between the area of the free space circle and its approximation in the cellular space. Each space has its own asymptotic ratio. The point of the comparison is to evaluate the

grids that also have a good fit to the “one unit step” from center cell to center cell. The square and rhomboid used above are the ones discussed. The hexagonal grid used is the one with a flat to flat diameter of s , which is the one that gives a center cell to center cell transition of s . At long ranges, the rhomboid approach ends up converging on an approximation of a circle somewhat better than the square, but not quite as good as the hexagonal.

3.2 Interfacing Squares and Hexes

Based on the above analysis, the rhomboid grid is *almost* as good as the hexagonal in several ways, so why not just use the hexagonal grid? The answer comes in trying to provide a reasonable interface between the two grids. A regular hexagon will either have at least one side length or one of the diameters that is an irrational value, which means only certain integer multiple numbers of rows and columns of hexagons will make a good fit with a square grid. Regardless of where those convergence points are for any two particular schema, there will still be significant loss where they are incompatible. Additionally, in the interim between the interfaces, groups of real number interpolations will have to be calculated both within one grid or the other and always to cross the interface. The advantage of a gridded system is the ability to quickly calculate transitions using counting and whole number math and systematic, repeating formulas, especially where this enables calculating relationships between non-adjacent cells without having to calculate states for intervening cells.

Because of the alignment of the rhombic dimensions with the square grid, there are a number of ways to interface between the two, creating a hybrid cellular space. A schema is proposed in the diagram below that provides a number of highly desirable characteristics for the composite space, and also illustrates a few of the challenges.

3.3 Integration

Figure 4 is constructed to illustrate the details of cell transitions in the hybrid space. The cells are numbered from 0 through B horizontally and 0 through D vertically. The interfacing shapes show that the two grids can be connected through a regular repeated pattern as the size of the interfaced region grows.

At first glance, the diagram shows that the *area* of the square grid can be maintained (that is, among the square cells, rhombic cells, and interface cells, all the cell coordinates that would have been in a homogeneous square grid of the same size are accounted for). This property scales down to a minimum of one hexagon (three rhomboids) and scales up without bound.

The coding scheme highlights one of the obvious challenges. The interface includes: regular square cells, regular rhomboid cells, truncated square cells, and augmented rhomboid cells. Some regular rhomboid cells that do not fit the repeated hexagonal pattern have been added (on the left side of the interface) to make the interface more efficient. The patterns mark the different transition algorithm data needed to move from one cell to another. Similar groups of patterns mark similar algorithmic placement of cells. A color coded scheme is provided at the end of the paper.

For the truncated cells, a square cell that overlaps with a rhombic cell is considered to be only the part of the square cell that is exclusive of a rhombic cell. Conceptually parallel to the idea of being truncated, some of the truncated cells (the “points down” ones) only share a side with three different cells. The effect on implementation is that change of position in two different directions could lead to the same place.

For the augmented cells, a square cell that overlaps with a rhombic cell is subsumed in the rhombic cell. In a similar conceptual parallel, these cells share a side with more than four other cells, and may share multiple sides with one other cell or have single geometric sides that border two cells. This can be handled in two primary ways: (1) ignore one of the border cells so there is no legal cell to cell transition where there is a physical sense that there should be one, (2) add custom decision and management criteria for transitions from these cells.

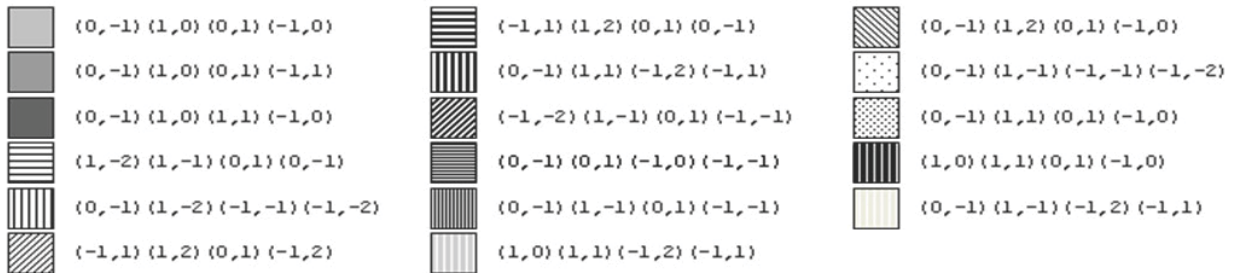
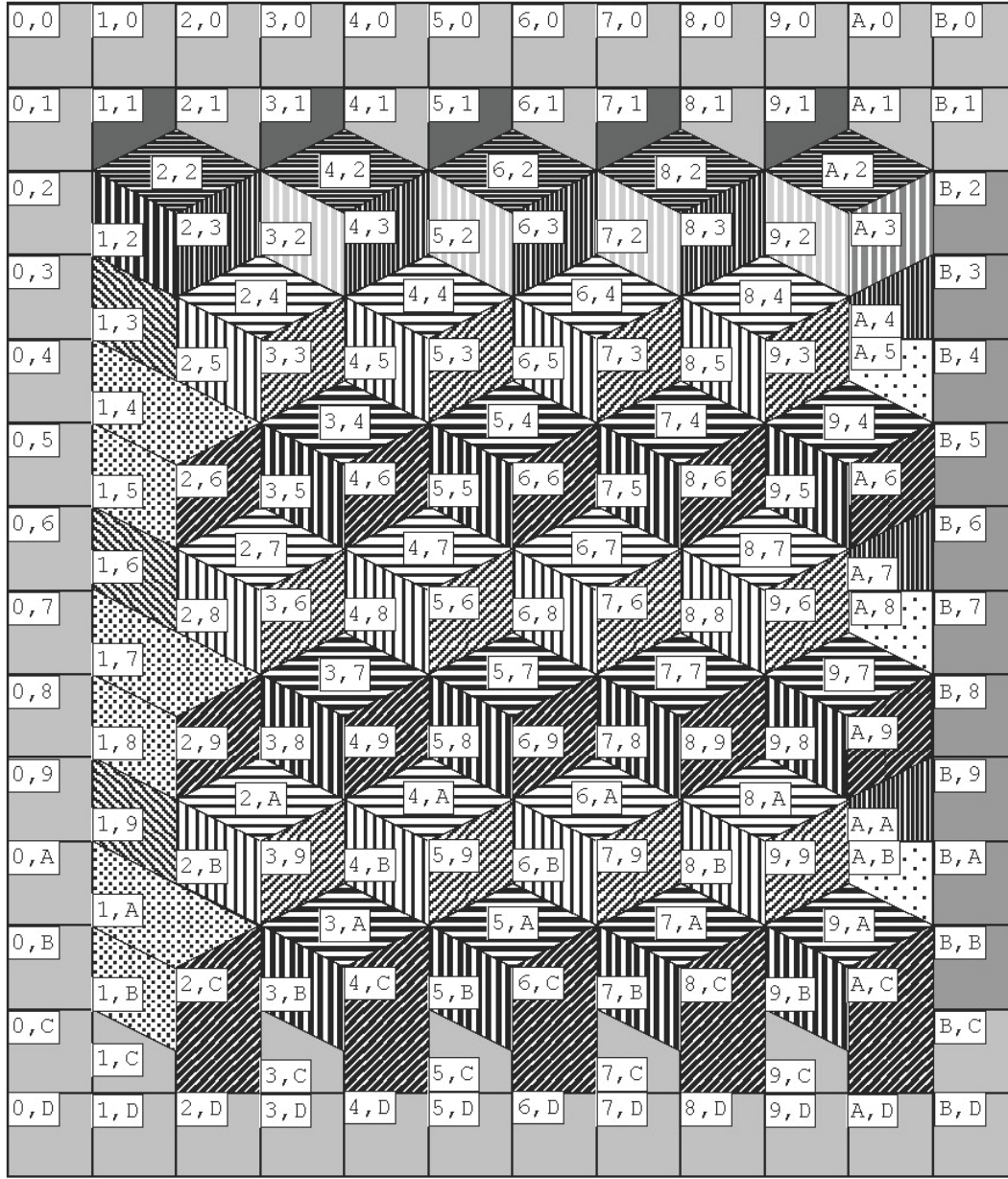


Figure 4. An Interfacing Schema with Transition Maps

The added cells are less problematic. They are simply additional rhombic cells that don't align with the existing pattern that are created to aid in the interfacing process.

Across all the interface cells, there are challenges with numbering and transitions, but these are relatively easy to deal with through the repetitive application of integer based algorithms. Their most significant challenge, however, is that they cause noticeable problems with the *flow*, of the grid.

3.4 Flow

As was earlier mentioned, the entire set of expected coordinates from the square grid are mappable to the hybrid grid, with a small number of difficulties and a number of reasonably easily handled special cases (This is especially true when you consider the typical application stated for this approach – the meshing of relatively large areas of the two systems to create a combined urban-open area that is tactically useful for wargaming. The number of normal cells grows geometrically with expansion of the area, while the number of interface cells grows only linearly, with the problem cells being a relatively small portion of those.) But the piecewise changes from cell to cell are not as important as the overall performance, especially over large distances and across multiple transitions – the flow of the hybrid grid. Using the base square grid as a referent, we will look at the overall horizontal and vertical flows.

The horizontal flow is the easiest to address and the simplest to evaluate. One can transition from the (1,x) to the eighth column cells in seven steps, the same as the square grid. Almost all of them are through a path that stays in the same row and most of them can be achieved with a series of moves to the right, again the same as the square grid. Since the variations are small in number and size, this seems to be a very good fit. Again, going back to the typical use case, the existence of a few places

where you have to move a little to the side to cross an area of rough ground does not seem to be incongruous.

The vertical flow has more challenges. The only paths through the rhombic grid that can maintain the number of crossing steps that the square grid has are ones that start near the vertical edges of the interface area. The minimum transit paths vertically through the center of the rhombic grid only cost about one more transition in ten, which is compatible with the distance loss of the squashed hexagonal grid, and isn't especially incongruous with the typical use case. Additionally, all the minimum vertical paths that don't run along the interface edge require some lateral movement. A large variation is not required, but it does affect a large number of paths.

When you couple the horizontal and vertical flow behaviors, an interesting characteristic emerges – there is an “easy” direction to transit the area, and a “hard” direction, that is, horizontal transition consistently requires less variation from the square grid performance than vertical. This behavior is not incongruous with many open, natural areas which have bias in the sloping and variation of the terrain, but it is not strongly compatible with all types of typical use case terrain (such as a groomed and manicured park in an urban area).

3.5 Additional Effects

When we note that this is only a function of horizontal and vertical as displayed in the example and not all grids in general, we see yet another interesting emergent behavior. Rotational symmetry of the grid allows us to swap the vertical and horizontal of the sample without changing the performance parameters relative to the original horizontal and vertical – the easy way becomes the hard way, and vice versa. This property creates the opportunity to, within the same overall map, have subsections of rhomboid grid that have different *orientations*, which

could be managed to impact that tactical characteristics of the terrain.

One additional emergent behavior of the approach is that swapping out top for bottom rhomboid configurations in the superposed hexagonal cells does not affect the long overall performance of interfaced gridded areas, but can cause additional local variations where the changes occur. Again, this behavior could be managed to support specific terrain effects that impact the tactical characteristics of the terrain.

As a final significant parameter of the hybrid grid, while it is not possible to easily integrate every possible combination of hexagonal and square natural areas, the interface is able to be changed in a regulated fashion at the edges in small increments. This means it would be relatively easy to integrate a wide variety of shapes for both types of sub grid.

4.0 CONCLUSION

In summary, the superposition of a rhombic grid over a base square grid provides the opportunity to leverage the advantages of both rectilinear and hexagonal wargame battlespaces and enables optimization for urban and natural terrain in a piecewise fashion with a minimum of compromises. This solution supports existing methods and properties of wargame entities and is compatible with existing AI templates and methodologies.

While the solution is interesting, and useful for both integration of legacy systems, what is more intriguing is the investigation into how we represent the warfighting environment, why we represent it in the ways we do, how we evaluate the characteristics of those representations, and how to leverage those characteristics to establish a trade-space across those representations.

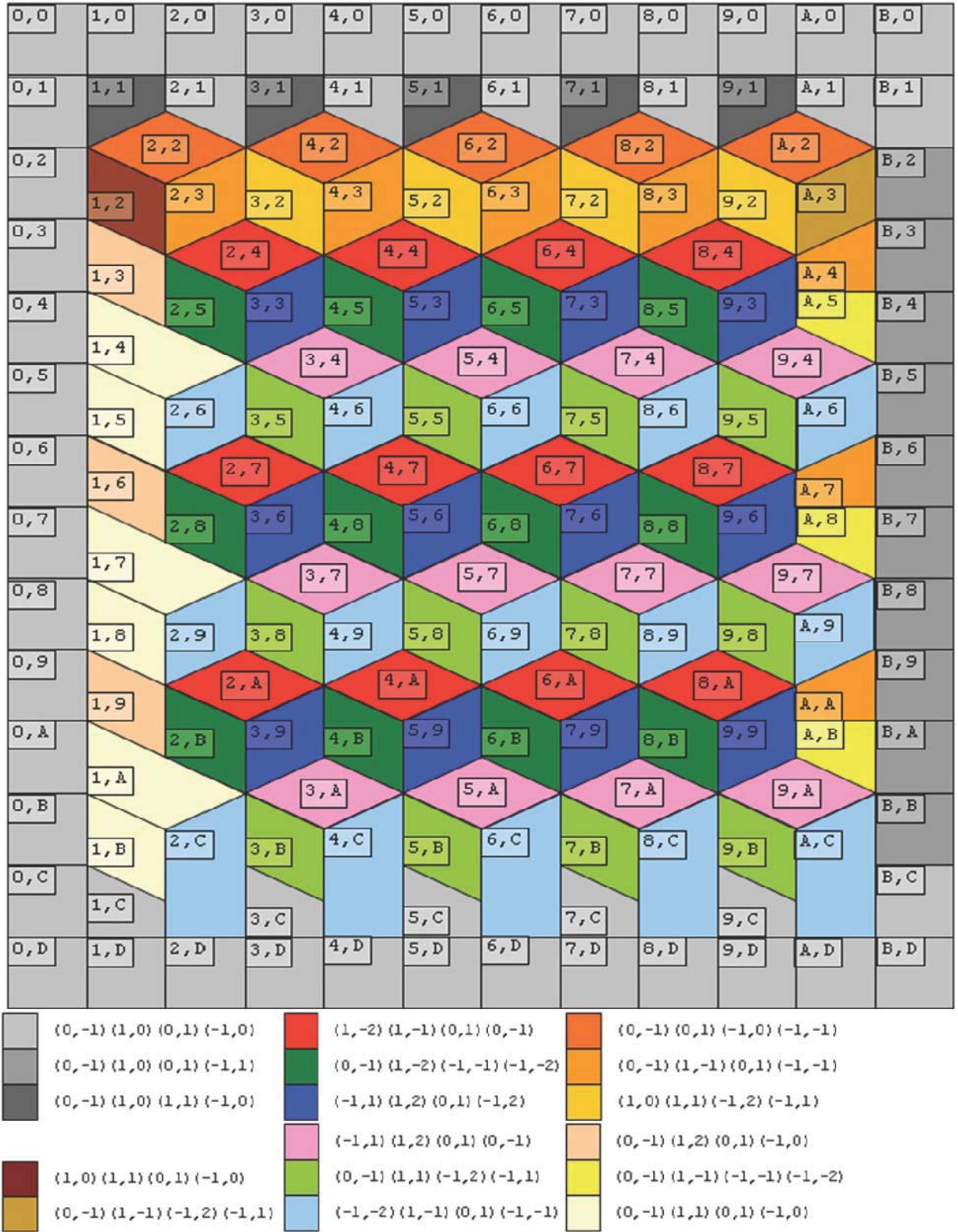


Figure 5. Color Coded Interfacing Schema with Transition Maps

1.4 Data Farming and the Exploration of Inter-agency, Inter-disciplinary, and International “What If?” Questions

Data Farming and the Exploration of Inter-agency, Inter-disciplinary, and International “What If?” Questions

Steve Anderson
Naval Surface Warfare Center
steven.e.anderson@navy.mil

Gary Horne
Naval Postgraduate School
gehorne@nps.edu

Ted Meyer
Naval Postgraduate School
temeyer@nps.edu

Larry Triola
Naval Surface Warfare Center
larry.triola@navy.mil

Data farming uses simulation modeling, high performance computing, and analysis to examine questions of interest with large possibility spaces. This methodology allows for the examination of whole landscapes of potential outcomes and provides the capability of executing enough experiments so that outliers might be captured and examined for insights. This capability may be quite informative when used to examine the plethora of “What If?” questions that result when examining potential scenarios that our forces may face in the uncertain world of the future. Many of these scenarios most certainly will be challenging and solutions may depend on interagency and international collaboration as well as the need for inter-disciplinary scientific inquiry preceding these events. In this paper we describe data farming and illustrate it in the context of application to questions inherent in military decision-making as we consider alternate future scenarios.

1.0 INTRODUCTION

What if energy sources became more scarce and suitable replacements are not cost effective? (Call this possibility: *Out of Gas*.)

What if renewable energy such as solar or wind became reasonably practical and widespread? (*Good and Green*.)

What if nuclear fusion became available to supply energy efficiently, safely, and at a fraction of the cost of current sources? (*Safe and Cheap*.)

Certainly these three possibilities are not the only energy futures that our world faces. And each of the three in and of itself poses challenges and opportunities for our military forces. But then also consider another very large global question we name *Coastal Flooding*: **What if** climate change factors

develop and result in an increase in the number and severity of calamities such as floods? Or **what if** it didn't? (*No Floods*).

Combining these two sets of what-if questions results in $3 \times 2 = 6$ possibilities already. One of which would be *Out of Gas* and *Coastal Flooding*, certainly a stark challenge. But even *Safe and Cheap* and *No Floods* might result in other challenges such as instability in former oil producing regions.

And with these six possibilities we have only begun to scratch the surface of the plethora of what-if questions that result when examining potential scenarios that our forces may face in the uncertain world of the future. Many of these scenarios most certainly will be challenging and solutions may depend on interagency and international collaboration as well as the

need for inter-disciplinary scientific inquiry preceding these events.

In this paper we describe a way forward to examine future possibilities using the methods of data farming. Data farming uses simulation modeling, high performance computing, and analysis to examine questions of interest with large possibility spaces. This methodology allows for the examination of whole landscapes of potential outcomes and provides the capability of executing enough experiments so that outliers might be captured and examined for insights.

1.1 What If?

The title of figure 1 provides the overarching philosophy of the What If? Network we are developing. All disciplines have strengths, but for the large challenges and broad scope of the questions we would like to grapple with, we need a truly multi-disciplinary approach.

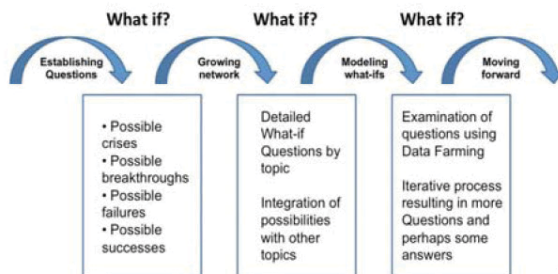


Figure 1. A Multi-disciplinary/ Multi-agency/ Multi-national What If? Network of Questions

And although we (the authors) work within the United States Department of Defense, these large challenges require a broadening of possible solution spaces that comes with looking across agencies. In our initial efforts we have focused on Department of Energy agencies. But certainly our network must grow beyond these two cylinders of excellence and indeed find ways of connecting the ideas found across agencies.

Finally, the challenges we face are global in nature and our collaborators in places such as Sweden and Finland have shown great

interest in the kinds of challenges that we as a world community share as well as great acumen in modeling, simulation, and data farming.

In figure 1 we have outlined three sets of what-ifs? The first block represents the big issues and our introduction gave you a brief look into the kinds of questions we have been establishing as a starting point. We will present our outline of a number of them in section 3.

The second block represents the detailed questions that will allow us to integrate possibilities on a level where insights and solutions might become clearer. We do not address specific questions in detail in this paper, although examples are given as part of the material in section 2. Also, in other work we are currently considering more specific and technology driven questions such as: What if a new craft was developed that would be a significant technological advancement over previous ship-to-shore transport capabilities? That development may have impact in the *Coastal Flooding* what-if mentioned earlier and this analysis leads to the additional question: what if this craft had a power supply stemming from the *Safe and Cheap* what-if?

Finally, the third block represents the data farming of the questions and we will give a very brief overview of data farming in the next section, although a deeper discussion can be found in our MODSIM World 2010 paper "Data Farming and Defense Applications" (Horne and Meyer 2010).

1.2 Data Farming Overview

Data farming is a collaborative and iterative process that requires input and participation by inter-disciplinary teams to be most effective (Horne 1997). It allows for an examination of a more complete landscape of outputs rather than one particular answer. Data farming also allows for the discovery of outliers that may be even more instructive than any general patterns that are discovered (Horne and Meyer 2005).

Data farming has been described as including six domains. The six domains make up the focus of the six sub-groups in a NATO Modeling and Simulation Group effort (NMSG-088) called "Data Farming Support to NATO" that is now in its second year of existence. MSG-088 members are performing two case studies, one in the area of humanitarian assistance / disaster relief and the other in the area of force protection. The MSG-088 subgroups are in the process of defining the six domains in detail as they apply to data farming, and the domains were described in our previous MODSIM Paper that we mentioned, but they are listed here for reference.

- Model Development
- High Performance Computing
- Rapid Prototyping of Scenarios
- Analysis and Visualization of simulation output
- Design of Experiments
- Collaborative processes

We believe all of the domains of data farming listed above will be important in the examination of the what-if questions we are considering. However, the domain of analysis and visualization promises to be a key domain and in the next section we explain why and give some detailed information regarding our proposed approaches in this domain.

2.0 WHAT IF? ANALYSIS AND VISUALIZATION

In general, analysis, and visualization in particular, have two broad purposes: 1) answering questions and 2) determining what questions to ask. Answering specific questions is accomplished by a traditional and ever-growing suite of statistical and graphical techniques, both well known (e.g., averages, regression, line plots) and more exotic (e.g., Yuen's modified t-tests, Kalman Filtering, trilinear graphs).

Often times, though, an analyst may be involved in a what-if process that is open-ended... the data being examined is

complex, voluminous and not well understood. Before well-defined questions can be asked, "exploratory" analysis of the data may be undertaken to establish an understanding of the data "landscape."

The class of analyses that are to be undertaken by the proposed What-If? Network requires that collaborative multi-disciplinary teams undertake exploratory-type analytic processes. These multidisciplinary teams of modelers, decision-makers, analysts, and subject-matter experts are what will form the What-If? Network.

These teams are not looking to simply provide basic summary statistics and trends from models. The intent is for these teams: to exercise a variety of scenario development models (e.g., subject-matter judgment, agent-based simulation, scenario network mapping, role playing, etc.); to data farm these resultant models; to integrate the models' results into a coherent landscape of potential outcomes; and to explore the landscape to gain an understanding of alternative futures. Individual modeling processes, undertaken by teams with limited disciplinary and collaborative breadth, are unlikely to provide results with verisimilitude that covers the breadth of real-world potentialities.

Real-world scenarios encompass outliers, second and third order effects and unintended consequences, runaway trends from feedback loops and dampening effects, emergent behaviors, social and human response, and nonlinearities and chaos. To generate potential real-world alternative futures multiple abstract, computational and human-based modeling techniques are required to move past the limitations of single classes of models.

Data farming a single simple model often results in analytic challenges in examining potential voluminous results. Exploratory visualization techniques offer capabilities in exploring landscapes of potential results. Integrating and exploring the results of

multiple models becomes a greater challenge.

Perhaps just as important as exploring the results of modeling efforts is the exploration of HOW and WHY the results are attained. Some modeling methods allow analysts to examine the state of the system over time, the evolution of the system, the interaction and behaviors of its components, and the changes to its inherent networks. This need to examine the “how” and “why” significantly increases the data to be examined beyond the data volume impact of data farming.

2.1 Visualization Methods and Examples

Visualization can play two roles in accessing this potential mountain of data that faces what-if analysts: aiding in the discovery of insights in the integrated whole of what-if scenario studies and in the compelling presentation of these results to policy makers. Our current focus is on exploration of the alternative future space, as finding the important insights is a more difficult problem.

Three exploration tools are powerful techniques to optimize the examination of high-dimension, high-volume data sets (Buja, McDonald, Michalak, and Stuetzle 1991):

1) Linked Displays

Typically information graphics provide a view of low dimensional data in easily understood representation. For example, histograms present univariate data, simple scatter plots represent bivariate data, and color and size can be added to a scatter plot to provide a third and fourth dimension. Linked displays tie two or more representations of the same data across multiple graphic displays. Each display can add to the overall dimensionality being presented at a single time. In the case of simulation, a playback of data can be tied to a representation of network state or other performance metrics.

2) Variable Focus

A data display can be adjusted in various ways to change the perspective being used to represent the data. In the simplest form the position and scale can be adjusted to zoom into detailed features or to zoom out to an overview. More complex forms of focus can include three dimensional rotation, geographic projection, and axis and parameter selection.

3) Interactivity

User interactivity can be used to adjust linking and focus in order to explore the relationships of data, select which portions of the data to examine, or determine how some parameters impact metrics.

These three techniques in combination can be integrated into the what-if process to interactively iterate model results and model development to hone insight into potential future outcomes.

As an illustrative example (Koehler, Meyer, McLeod, Burke, Johnson, and Barry 2007), a changing “focus” can drill down into more detail and a better understanding of a scenario. In this example a time-stepped simulation of a combat scenario is data farmed over 50 replicates. Results may vary in detail:

- 1) **Single Numeric Statistical Summary Value in Text Form:** A model scenario results in 45 Blue casualties, on average, per run.
- 2) **Numeric Statistical Summary Values in Text Form:** the model indicates 45 Blue casualties, on average, per run; 65 Maximum Blue Casualties over all runs; 11 Minimum Blue Casualties over all runs.
- 3) **Time Series of Numeric Statistical Summary Value in Line Plot Form:** Figure 2a shows the average number of casualties each hour in 50 executions of the scenario.

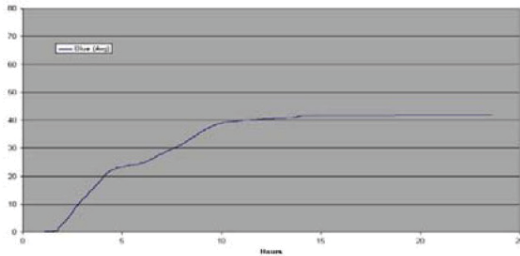


Figure 2a.

4) **Time Series of Numeric Statistical Summary Values in Line Plot Form:**

Figure 2b shows both the average as well as the maximum and minimum number of casualties for each hour in the 50 replicates.

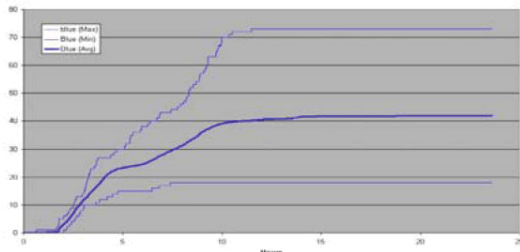


Figure 2b.

5) **Jitter Plot of ALL Values Providing Distribution Details:** Figure 2c shows the average, minimum and maximum, but it also shows a full distribution of all casualty values for each hour for all 50 replicates

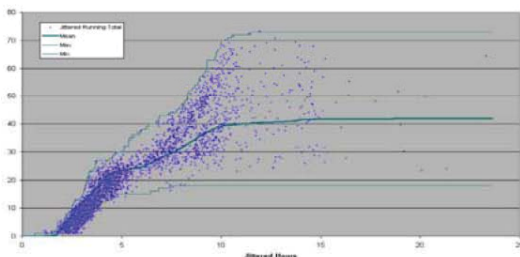


Figure 2c.

Figures 2 a to c demonstrate how more detail can reveal important information. A close examination of the distribution of casualties in Figure 2c reveals a bimodal distribution that occurs at about 7.5 hours into the scenario. An examination of this time frame shows a bifurcation event in the scenario that results in a split in the results. Often, analysis techniques will *reduce* the amount of available information to the

analyst or decision maker. Visualization provides methods for increasing the available information.

Another example of using visualization techniques to increase the display of information has been experimented with for sometime. Density Playback is a technique for examining multiple replicates or scenarios in an overlaid fashion to highlight similarities or differences in the model.

In density playback, data of interest is plotted in whatever state space is desired using scaled transparency. The amount of transparency is dependent on the number of model executions. An obvious example of this technique is represented in the “Death Star” scenario represented in Figure 3a and 3b where the spatial position of agents are represented using density playback.

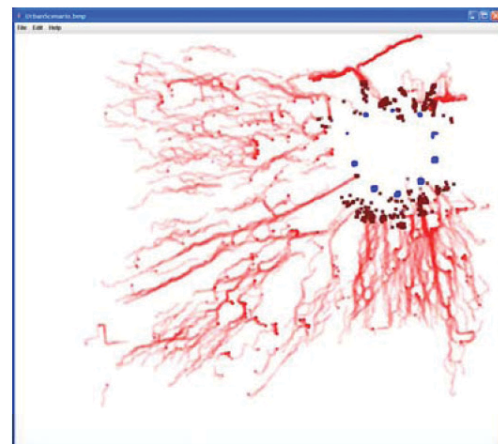


Figure 3a. Density Playback – Random

The scenario evokes Luke Skywalker’s run at the “Death Star” in the 1978 Star Wars movie. The scenario has a central target that is extremely well guarded by a ring of well-armed blue agents in the top right of the figures. Blue does not move and is well positioned. Fifty unarmed Red agents only need to penetrate the ring to win the scenario. 26000 runs of this scenario resulted in less than 100 Red wins.

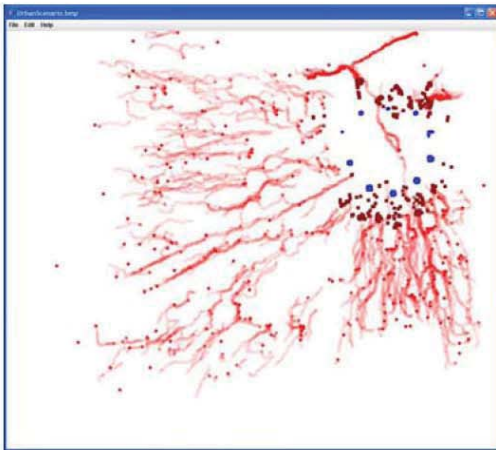


Figure 3b. Density Playback – Red Wins

Figure 3a is a random selection of 50 executions of the scenario from the 26000 executed displayed at timestep 350. Transparency is used to display the 50 red agents' trails. Dark trails represent locations where many agents traveled over multiple executions. Note that no agents penetrate the ring.

Figure 3b represents 50 executions selected from the small set of Red wins. Note that only two very specific paths led to Red winning. This example shows that changing focus to filter only on data associated with an outlier result results in a highlighted display of the path Red needs to take to victory.

The potential scenarios that are to be addressed by the What-If? Network are far more complex than the simplified examples shown. Additionally, the "terrain" used by these potential scenarios are likely to be abstractions of a political/social/economic landscape rather than the simple spatial examples shown. The techniques shown can be applied to the more complex cases and can be further explored using linked displays, interfaces that provide interactivity, and interfaces that link results to specific scenarios and replicates.

The techniques to be used by the What-If? Network will range from simple descriptive statistics to sophisticated machine learning algorithms. All analytic techniques have

strengths and weaknesses, as do all modeling and techniques. And, of course, the combination of methods to best understand the potential outcomes will depend on the questions at hand. Thus, at this point we will now turn to describing some of the question areas to be considered by our growing What-If? Network.

3.0 THE QUEST FOR WIN-WIN SOLUTIONS

"The trouble with military force structure is that it typically outlives the *geopolitical context* that called it into existence."¹ Whether the U.S. government faces future fiscal constraints, or makes changes to its overseas commitments, the U.S. Department of Defense may choose to reevaluate its force structure. Deer hunters can attest that the most effective hunters will aim at where the deer will be, rather than where the deer was. Leading the target is as important to DoD force planning as it is to hunting deer. Getting future *geopolitical contexts* right, more or less, is an important part of any future debate.

The study of future *geopolitical contexts* is an obscure and delicate form of defense analysis that requires evaluating interdisciplinary trends and data, tracking science and technology investment areas, monitoring acquisition programs, conceptualizing current and emerging operational needs and missions, and translating all of this into robust recommendations for future concepts of operation, development portfolios and force structure.

Whereas most military planning scenarios are focused upon political-military actions and reactions of nation-states, the second half of the Post Cold War Era is replete with evidence and examples of trends, actions, and events that have shaped the international stage, and yet are neither

¹ This observation is attributed to retired defense analyst, Mr. James S. O'Brasky.

initiated by, nor responded to by nation-states. A set of highly informative future contexts arise from examining certain global mega trends and wild cards such as climate change, wild card disasters, wild card revolutionary technology developments, global resource limitations and other trends. National security relevant illumination comes by relating the geopolitics and U.S. mechanisms of planning to address such events.

As civilization enters the second decade of the 21st century it is confronted with an abundance of diverse pressures the sum of which seems to outweigh each as individual problems. The United States and the U.S. Department of Defense must develop methodologies to enhance the positive synergies and mitigate the negative synergies of this collision of crises. Terrorism, tribalism, fossil fuel shortages, water resource shortages, challenges in health care and education, new technologies both helping and threatening, economic challenges, globalism and the flat-world syndrome, disruption of old alliances and the formation of new ones, climate change, mass migrations, resource limitations on numerous fronts and failures of governance, policies and leadership and many other issues wash wave upon wave in cascading calamities of white water of our changing times. Halal and Marien have called this situation the "Global MegaCrisis". While the authors of this paper work for the Department of the Navy within the U.S. Department of Defense, the issues we uncover consistently require coherent and integrated whole-of-government response plans, many of which are time-critical in nature (Halal and Marien 2011).

Looking at each issue individually and trying to develop individual solutions almost ensures that some solutions will interfere with others or that solutions will simply be too many and too costly to implement. What is here proposed is to construct a framework for the various crises such that they may be addressed in a coordinated,

cross-domain, cross-disciplinary approach that leverages the emerging capabilities of data farming.

3.1 The World MegaCrisis Framework

The World MegaCrisis Framework (see the appendix) contains a set of ten categories of challenges with a list of more detailed topics and sub-topics listed under each.

1. Global Physical and Biological Dynamics such as climate change, weather and geological disasters, biodiversity loss and biological change in habitat
2. The limitations of natural and perhaps man-made resources and the limitations on or help to civilization which emerge from such resources and their limitations
3. Changing of demographics and populations such as the radical drop in Russian population and native European population and the rise of population in Europe from the Islamic countries stretching from Tunisia to Indonesia
4. How societies within nations and regions view themselves and choose or not to act as coherent groups to include the range of effects from socialized democracy to genocide
5. Economics and manufacturing issues include currency, national and international debt, trade imbalances, ability or lack thereof to manufacture goods
6. Knowledge includes educating the young to function in society but also includes the context of information or misinformation which drives public opinion and decisions
7. Infrastructure covers the range from electrical power grids, to water and sewage, to health care facility, to educational institutions, to transportation

8. Transportation is also suggested as its own topic to cover methods and the societal, logistical, and transnational implications
9. Technological breakthroughs can destroy cities or provide them endless electricity and much more depending on what they are and how they are used and many potential world-changing wild cards exist in this category
10. Leadership and governance at the human level and the institutional and legal levels can either be facilitation to or inhibitors of solving crises

This is not an exhaustive set, yet all of which have significant worldwide impacts. However, note that some challenges are supersets of others and that challenges listed under one heading often offer the potential to exacerbate or mitigate challenges listed under another heading. These interdependencies are a critical feature of this framework. As an example thread, consider that Climate Change shown under the heading of Global-Physical / Biological Dynamics leads to such things as Water Cycle Changes and Biodiversity Loss and impacts the availability of Resources such as Arable Farm Land. Fossil Fuel use shown under Resources Limitations is seen by many including the International Panel on Climate Change to directly impact Climate Change and provides a feedback loop for Resource loss. Under Technological and Scientific Breakthroughs alternative energy options and new material options create Resources and perhaps present new Resource limitation challenges.

The topics listed in the MegaCrisis Framework, each have self-organized communities of interest. Community cohesion can vary from topic to topic. In some instances, a given community may have significant shared outlooks and beliefs, while other communities are highly polarized on central points. To remain objective, data farming must focus on

“possibilities” rather than probabilities and predictions.

Many threads can be drawn from populating a database rich in diverse contexts such as those suggested by the MegaCrisis Framework. These can be perhaps too complex and interwoven for simple brute force human interpretation. Thus, automated methods to mine this data and to develop additional information from the cross-connecting of the threads are essential. Looking through only one lens will not provide realistic nor optimum approaches for addressing the many challenges. Connecting the dots may in fact provide a clear synergistic set of solutions that may be balanced and implementable. It is hoped that resources will be made available to farm this data and connect the dots as well as to identify and flesh out other contexts that can help governments to successfully confront this tumult of challenges, and thrive.

Single-issue agendas are often fraught with unintended consequences. The so-called “Biofuel Controversy” is an example of ethanol producers who seek to increase the availability of ethanol to consumers at a reasonable cost (Hazell and Pachauri 2006). While this is an admirable single-issue agenda, there are scores of documented examples around the world where arable land that is used to generate crops for biofuels is not available to produce food – creating a situation where biofuel production causes famine. Biofuel production that leads to the starvation of impoverished peoples is a “win-lose” situation, at best (Faissner 2010).

If you are a senior decision maker, it almost does not matter where you sit, in any government in the world (United Nations 2007). You will be faced with the growing demand to make highly informed cross-disciplinary cross-agency decisions that invariably reach outside of any given historic stovepipe. Ideologically, the goal is to make coherent and integrated decisions that

foster far reaching benefits, and minimize adverse consequences. The quest is to achieve “win-win” situations wherever possible. To do this will require new patterns of thinking, new questions, new methodologies and techniques, more data, and new and better tools.

4.0 SUMMARY

Pragmatically, how do we achieve these ends? There are some mandates, all of which are important as we implement data farming practices, techniques and approaches:

- We must get to the point where we are asking the right questions.
- The path to the right questions will be iterative.
- We must resist the temptation to delete, discount, or discard “outlier” data and trends.
- We will embrace audacity, integrity, and humility.
- While top-down approaches would be very useful, we will start with the data and tools we already have in hand. As top-down, middle-out, bottom-up strategies interact and normalize, we will need new data and new tools to effectively answer emergent questions.
- We have a sense of urgency. Thoroughly exhaustive approaches may not be practical or ultimately useful to solve near-term challenges.

But, we are still just beginning to ask the detailed what-if questions in the second block of figure 1 and only setting ourselves up for the modeling, simulation, and data farming efforts at this point. And that is why we are coming to MODSIM World now with this work: because inter-agency, inter-disciplinary, and international expertise will

be needed as we move to the third block in figure 1. Thus we would like to invite you, members of the modeling & simulation community, to join our What-If? Network and contribute to the quest for win-win solutions.

5.0 ACKNOWLEDGMENT

The authors would like to thank the members of the growing inter-agency, interdisciplinary, and international network of people who contributed their ideas in the quest for discovering win-win solutions through data farming.

6.0 REFERENCES

- Buja, A., McDonald, J. A., Michalak, J., Stuetzle, W. 1991. Interactive Data Visualization Using Focusing and Linking. *Interactive Visualization*. Pp. 156-163.
- Faissner, K. 2010. *Biofuel Causes Famine*, Current Concerns, No 7, April 2010.
- Halal, W. and Marien, M. *Global MegaCrisis*, The Futurist, May-June 2011.
- Hazell, P. and Pachauri, R. eds. 2006. *Bioenergy and Agriculture Promises and Challenges*, 2020 Focus No. 14, IFPRI/TERI, Washington, DC/New Delhi.
- Horne, G. 1997. *Data Farming: A Meta-Technique for Research in the 21st Century*, briefing presented at the Naval War College, Newport, RI.
- Horne, G. and Meyer, T. 2005. *Data Farming Architecture*. Proceedings of the 2005 Winter Simulation Conference, eds. M. E. Kuhl, N. M. Steiger, F.B. Armstrong, and J. A. Joines, 1082-1087. Orlando, FL.
- Horne, G. and Meyer, T. 2010. *Data Farming and Defense Applications*, MODSIM World Conference and Expo, Hampton Roads Convention Center, Hampton, VA, USA 13-15 October 2010.
- Koehler, M., Meyer, T., McLeod, A., Burke, C., Johnson, S., and Barry, P. 2007. *Visualization of Systems of Systems Simulations: Density Plots and Trajectory Storyboarding*. Systems Conference 2007 1st Annual IEEE.
- United Nations. 2007. *UN Energy, Sustainable Bioenergy: A Framework for Decision Makers*. New York.

Appendix

The World Mega Crisis--Collision of Crises									
Global-Physical/ Biological Dynamics	Limited/Limiting Resources	Demographics Dynamics and Imbalances	Fractures in Societal Cultural Identities and "Self Image"	Economics and Manufacturing Dynamics	Knowledge - the Maintenance, Control Acquisition thereof	Infrastructure Issues	Technological/ Scientific Breakthroughs	Transportation System Dynamics	Failures/Challenges in National and International Governance, Policy and Migration / Immigration Policy
Climate Change	Rare Earths	Native gain of population from reproduction	Tribes / Tribalism	International Economic agreements - debtors/creditors	Fourth Estate Failure	Failing Infrastructures	Human Brain Reverse Engineering and understanding- artificial minds	Failure of Transportation systems	Migration / Immigration Policy
Solar Activity/Electric Grid Failure	Arable Crop Land	Loss of national populations from decreased fertility rates / disease healthcare	Nations and nationalism	Trade imbalances	Education and Training --- Costs and Processes and cross-cultural differences	Rising and Improving Infrastructures	Ubiquitous and Revolutionary Computing--infinite memory and processing speed available on everything	Expensive Transportation	Economic and Trade Zones and Favored Nation Candidates
Volcano Eruptions	Fresh Water Sources	Migrations with positive forcing functions	Religions---Forces toward fundamentalism and exclusivity or inclusiveness and openness	Manufacturing Embalances	Good and Bad Propaganda	Military Capabilities and Asset Reaset needs and methods	Instant, high fidelity communications	"Instant" World Wide transportation	Nuclear Proliferation Treaties
Tsunamis	Trees/ Rain Forest/ Timber	Migrations with negative forcing functions	"Culture" and societal common visions	Rich -Middle Class -poor dynamics	Keeping an accurately informed electorate		Nanotechnology comes of age- structural graphene, medicinal use, etc.		NATO and United Nations Authority, Responsibility and Resourcing
Violent Weather	Easy access metals	Population age balance	Formal societal "agreements" -generation to generation/ employer to employee/ government to governed	Fiat-worldism and gapology	Instant communications impact on freedom and freedom repression		Metamaterials --potential revolutions in energy computing, communications, stealth, construction		International Law and Local Governance
Biodiversity loss	Fossil Fuels	Population sex balance	Intensity and Diversity of Value systems	Market Impacts and manipulations	Information Control and Spin by Elected Governments, businesses, "interest groups", criminals, tyrants		Deep Electrical Storage		Ungoverned / Undergoverned /Misgoverned nations and regions
Disease evolution/ pandemics	Radioactive/ useful elements	Human evolutionary trends		Currency Exchange and Primary World Currency Choice	Digital Electronic Learning Opportunities		Wireless Energy Transfer		Tyrannies, dictators and U.S. responsibilities
Water Cycle Changes	Navigable Water			Resource Competition	Virtual Reality Education, Idea Exchange and Mind- Enhancement		High Fidelity, Complex-Part- System, Multi-Material 3-D "Printing" in high volume manufacturing regime		Personal freedom vs security in the U.S. and internationally
	Crop Types						Breakthrough in Stealth --either ending it or perfecting it		International Infrastructure and
	Crop quantities						Nuclear Fusion		Geoengeering projects War, Sanctions, and International Behaviors of National and non-nation actors
	Fertilizer sources						- Low Energy Nuclear Reactions		
	Wind Resource								
	Water Current-Flow								
	Resource								
	Geothermal resources								
	Ocean Food Chain Basis								

1.5 High Resolution Visualization Applied to Future Heavy Airlift Concept Development and Evaluation

High Resolution Visualization Applied to Future Heavy Airlift Concept Development and Evaluation

A. B. (Ford) Cook
MECHSIM Incorporated
ford.cook@mechsim.com

T. King
MECHSIM Incorporated

Abstract. This paper explores the use of high resolution 3D visualization tools for exploring the feasibility and advantages of future military cargo airlift concepts and evaluating compatibility with existing and future payload requirements. Realistic 3D graphic representations of future airlifters are immersed in rich, supporting environments to demonstrate concepts of operations to key personnel for evaluation, feedback, and development of critical joint support. Accurate concept visualizations are reviewed by commanders, platform developers, loadmasters, soldiers, scientists, engineers, and key principal decision makers at various stages of development. The insight gained through the review of these physically and operationally realistic visualizations is essential to refining design concepts to meet competing requirements in a fiscally conservative defense finance environment. In addition, highly accurate 3D geometric models of existing and evolving large military vehicles are loaded into existing and proposed aircraft cargo bays. In this virtual aircraft test-loading environment, materiel developers, engineers, managers, and soldiers can realistically evaluate the compatibility of current and next-generation airlifters with proposed cargo.

1.0 INTRODUCTION

The application of common 3D visualization tools (Computer Aided Design, Computer Aided Engineering and Computer Graphic Animation (CAD/CAE/CGA)) along with a depth of knowledge in military vehicle transport requirements, capabilities and policies can be very fruitful in the development and evaluation of future heavy-airlift concepts.

Unbeknownst to the casual observer, the cargo carrying capacity of military heavy airlift cargo planes often defines the size limitations for military weapon platforms [1]. For example, the Stryker family of eight-wheeled vehicles, which is the backbone of the US military's Interim Force, was acquired with a C-130 transport requirement that limits the vehicle external dimensions, overall weight and center of gravity, as well

as maximum axle loads. Another major acquisition program, the Future Combat System (FCS), which, until recently, was to be the primary vehicle platform for the Objective Force, initially had a C-130 requirement as well. But due to competing requirements, including ballistic protection, this requirement was dropped in favor of a less restrictive transport requirement: three FCS vehicles in a C-17 aircraft. This new requirement still imposed significant size restrictions that subsequently redefined the size and weight boundaries for the program. Yet another major acquisition program confirming this trend is the Joint Light Tactical Vehicle (JLTV), with a C-130 transport requirement that constrains the JLTV family of vehicle's size and weight.

This size-defining mechanism works in both directions. Future heavy-airlift platforms are defined and evaluated by the vehicles and cargo that they can carry. A future airlifter

that is capable of delivering heavy vehicles, one at a time, or in multiples, is obviously more attractive than a less capable platform. In fact, formal government requests for information from industry regarding future heavy airlift concepts are usually defined in terms of the vehicles they must be able to transport. However, total airlift capacity is obviously not the only objective of a heavy airlift concept [2].

Future heavy airlift concepts are evaluated against a variety of performance factors beyond dimensional, structural, and weight carrying capacities. These additional factors include range and speed of delivery, fuel efficiency, takeoff and landing capabilities (runway requirements: length, condition, elevation), self-protection capabilities, and cargo handling capabilities (ease and speed of loading, securing, unloading) among others [3]. The effectiveness of a future heavy airlifter platform development program depends upon clear communication of the capabilities inherent in a design, the ultimate impact of these capabilities, as well as critical feedback from a wide audience of critical stakeholders. Generally speaking, these stakeholders include buyers, builders, users and the general public. More specifically, these stakeholders include congress and the media, senior level military commanders and civil service leadership, program managers and engineering staff, cargo handling personnel (including Air Force Loadmasters), and soldiers among others. Each of these stakeholders has valuable input and feedback that, given the opportunity, can be facilitated, solicited and obtained via a visually accurate concept demonstration visualization.

2.0 FUTURE HEAVY AIRLIFT CONCEPT VISUALIZATIONS

Future Heavy Airlift Concept Visualizations are technically oriented, highly-descriptive, and concise videos based on conceptual CAD geometry, supporting 3D graphic objects and environments, thorough knowledge of the concepts of operations, and realistic limitations. MECHSIM Defense Transportation Technologies, Incorporated (MECHSIM) has provided technically oriented concept demonstration visualizations to the military heavy lift transportation community since 2000. Concept demonstrations have been provided to US Army Defense Ammunition Logistics Activity (AMMOLOG) at Picatinny Arsenal, The Boeing Company Advanced Mobility Interface Systems group, The Boeing Company Phantom Works, and The US Army Aviation Applied Technology Directorate (AATD) at Fort Eustis, among others.

Effective concept demonstrations clearly and accurately convey the need for a new heavy airlift concept, the essence of the proposed concept, the technical features and capabilities, the concepts of operation, as well as the ultimate impact upon the target industries. Figure 1 shows still images from the Joint Heavy Lift (JHL) High Efficiency Tilt Rotor (HETR) concept visualization produced for AATD in 2010. These highlights illustrate a sampling of 1) the need for a new heavy airlifter to overcome infrastructure limitations, 2) the layout and capacity of the HETR concept, 3) the technologies being developed in support of JHL HETR, 4) the payload/range capabilities of JHL HETR, 5) the flexibility of the HETR design in takeoff and landing modes, and 6) the range of operations that JHL HETR can support including military operations, homeland security and humanitarian relief [4].

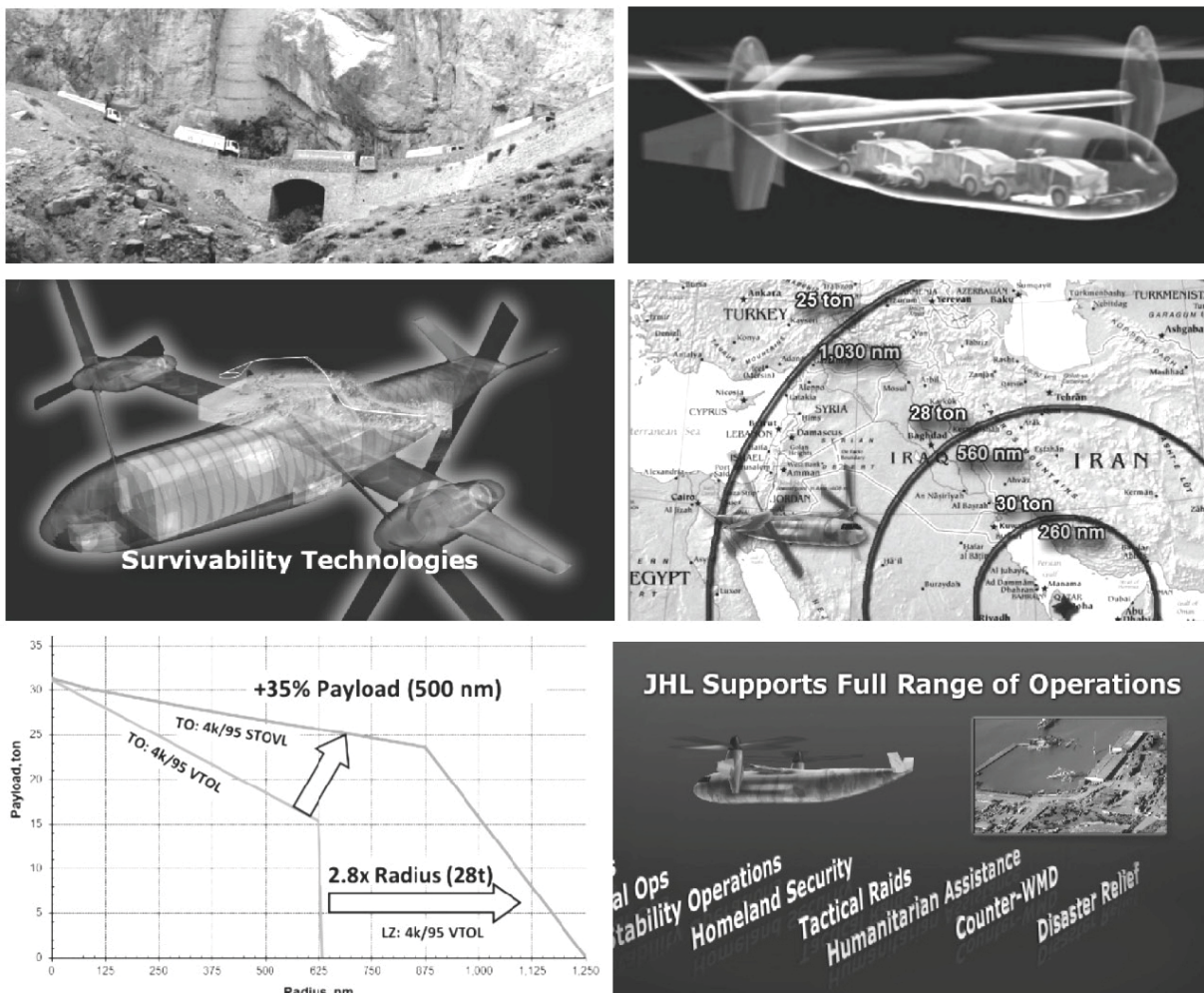


Figure 1. Joint Heavy Lift High Efficiency Tilt Rotor concept visualization still Images

2.1 Future Heavy Airlift Concept Visualizations: Process

Visualization development begins with interviews of the proponents and developers, along with background research to obtain a full understanding of the concept. Data collection ensues, including CAD model data and supporting media: brochures, presentations, media reports, live photographs and video clips. These are reviewed for usefulness in conveying the intended message. Next, a descriptive storyboard is formulated in close coordination with the client. Depending upon budgetary and schedule constraints, the concept demonstration may incorporate a narrative script with voiceover, sound

effects and background music. The storyboard continues the collaborative development and review process until all parties are satisfied that it conveys the intended message in a concise and easy to understand manner.

The geometry for the concept is obtained from the developers and converted into the animation software format and optimized for visual impact and rendering efficiency. This includes deletion of unnecessary features and addition of texture maps, colors and lighting effects. Supporting objects, including other vehicles, material handling equipment, cargo representations are obtained or created and prepared for animation and rendering. Supporting

environments are obtained or generated to illustrate concepts of operation in appropriate realistic scenarios. Once all geometry import and clean up is complete, it is adjusted for consistency in level of detail and artistic style and reviewed in detail with the client. Upon approval, computer generated graphic media production ensues as quickly as possible.

When the storyboard and script are finalized and substantial animation media is complete, composite video editing begins and draft video segments are produced and reviewed. This iterative process continues until a full draft concept demonstration video is produced. This draft video is reviewed at length with the client as well as other critical stake-holders. Upon successful modification, a complete concept demonstration video is delivered in multiple media formats, including standard definition, high definition and web-streaming video. In addition, if requested, the final video and associated clips are edited down to smaller segments targeted for specific applications including trade-show looping videos and videos targeting certain groups.

2.2 Future Heavy Airlift Concept Visualizations: Tools

While none of these activities alone comprise basic research, the balanced and skillful application of the collective toolset results in high impact cutting edge visualizations. The required skill-set includes: story-telling/crafting, communication, teamwork and problem solving, budget, schedule and resource allocation, familiarity with the industry, its language and challenges, animation and video editing, audio composition and voice-over talent, and attention to details. The ability to balance competing objectives, including client wishes, budget, schedule, employee resources and skills, is paramount. Visualization development also requires leading edge computer hardware and software, including multiple processors, large processing and graphics memory,

multiple monitors, and substantial hard-drive capacity and speed for rendering storage. A network of high-powered PCs can expedite the iterative rendering process. Software requirements include CAD geometry translation software, cutting edge commercial computer-graphic-animation software, video editing software and audio composition software.

2.3 Future Heavy Airlift Concept Visualizations: Application

The most obvious use of future heavy airlift concept visualization is in marketing venues. Quality visualization persuasively conveys the need for and utility of a new concept. However, a much less common, often overlooked and undervalued, application is in the development and evaluation of the new concept. New concepts typically are born out of discussions among end-users, technology developers, leadership and visionaries. Discussions evolve into sketches and bullet points. White papers are generated, and some preliminary quantitative analyses are often conducted. These are the seeds of a new concept. If a concept is believed to hold substantial merit, an organization may begin to apply engineering resources to its development. It is at this early, critical stage, that technical visualization tools can be leveraged to dramatically enhance the concept development process.

This innovative approach was effectively used by the US Army Armament Research, Development and Engineering Center (ARDEC) Program Manager for the Joint Modular Intermodal Distribution System (JMIDS), Mr. Doug Chesnulovitch, to expedite development and reduce costs. As shown in Fig. 2, JMIDS was a Joint Concept Technology Demonstration program, consisting of three components that improve the efficiency of heavy cargo distribution operations: 1) JMIP platforms (Joint Modular Intermodal Platform) that are PLS (Palletized Loading System) load handling capable, aircraft direct-load compatible, and

ISO container compatible, 2) Joint Modular Intermodal containers (JMIC) that are multimode compatible, and 3) electronic tracking devices. ARDEC's Ammunition Logistics Activity (AMMOLOG) worked with multiple developers, the user community, and joint leadership for a period of several years to develop a variety of concepts. These concepts, ranging from special pallets and flatracks to a variety of containers, eventually converged to the JMIDS system [5].



Figure 2. Joint Modular Intermodal Distribution System concept demonstration.

AMMOLOG utilized technical concept visualizations throughout the development process to capture and illustrate the features, capabilities and operations of each concept without building expensive physical prototypes. These visualizations were used in working group meetings and leadership reviews to rapidly and accurately convey concept information and solicit quality feedback. Upon viewing the visualizations, the audience members were able to immediately collaborate on the design and contribute valuable feedback from their unique and diverse perspectives. This saved immeasurable time and resources normally devoted to development, building and testing of hardware prototypes. Thus the virtual reality visualizations stimulate creative thinking, problem solving and brainstorming among groups. This same approach has been applied to several alternative heavy lift concepts from The Boeing Company as shown in Fig. 3.



Figure 3. Heavy lift concept visualization images created for Boeing (C-17B density loading, C-17B bridges capability gap, ultra heavy lift Hybrid Thermal Airship, Pulse Ejector Thrust Augmenter Pulse-Jet for VTOL).

3.0 FUTURE HEAVY AIRLIFT CAPACITY ANALYSIS

Future heavy airlift concepts can also be evaluated in virtual space for cargo carrying capacity, and compatibility with existing and planned cargo types. This can be done on a gross level with basic overall dimensional comparisons, and then with increasing fidelity as the design matures and engineering data becomes available. Compatibility assessments between large payloads and heavy airlifters can be conducted in both directions: A future airlift asset can be evaluated against existing heavy cargos, and a new weapon platform or vehicle, can be evaluated for compatibility with existing heavy airlift resources.

Figure 4, below, shows a virtual C-130 test loading aircraft. This model has the basic exterior geometry of a C-130, including props and landing gear, for visual effect. But more importantly, this model has detailed interior cargo bay geometry, and floor, ramp and cargo handling features that are validated for accuracy. The interior surfaces were developed from laser-radar scanning with a quarter-inch tolerance. The floor, including rollers, rails and tiedown points is based on published Air Force data, and has also been validated for accuracy. This model was used extensively in the development of the US Army Stryker Light Armored Vehicle.



Figure 4. Virtual C-130 test loading aircraft.

The Stryker program is a family of vehicles that require C-130 transport. This requirement challenged the vehicle developers in several areas. The vehicle had to meet certain C-130 weight criteria while also being operational upon delivery. In addition, the aircraft has certain structural limitations in the floor and ramp that limit axle loads both statically and dynamically during the loading process. Finally, the Air Force requires a crawl space either over or beside the vehicle for the loadmaster to move from front to rear of the aircraft during flight. MECHSIM modeled the physics of the suspension system together with the vehicle CAD geometry and conducted virtual aircraft test loadings to measure clearances (or interference), axle loads, and available crawl space. These analyses were conducted iteratively with incremental adjustments in suspension settings and design modifications until the vehicle was successfully loaded and secured in the virtual environment. This process expedited the progress towards the ultimate vehicle designs and loading procedures, and substantially reduced the cost and delays of physical testing that would have been required otherwise. Without this type of tool, the Stryker family of vehicles may have never received C-130 flight certification. Figure 5 shows a comparison of crawl space clearance in the model and in reality.

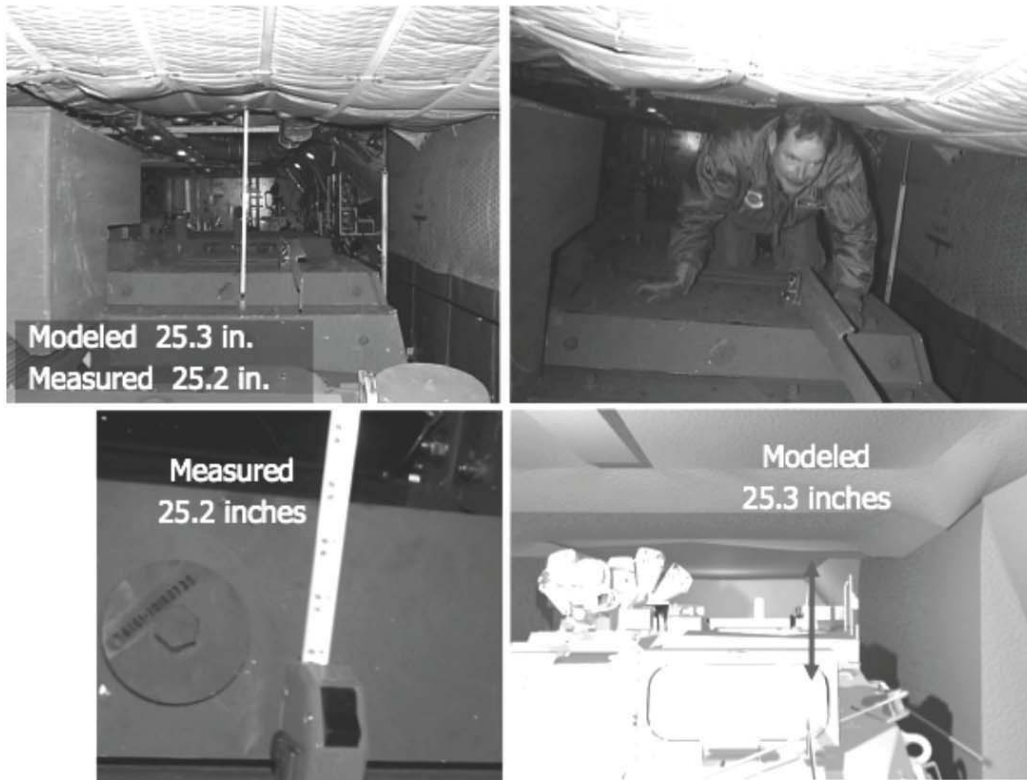


Figure 5. Comparison of Stryker/C-130 crawl space clearance in the model and in reality (crawl space between vehicle hatch and C-130 wingbox, crawl space live test, measured crawl space overhead clearance, and modeled crawl space clearance).

In addition to predicting clearances and axle loads, virtual aircraft test loadings are useful for developing load plans, including vehicle placement and chain tiedown patterns, for existing and developmental vehicles and aircraft concepts. This type of analysis was

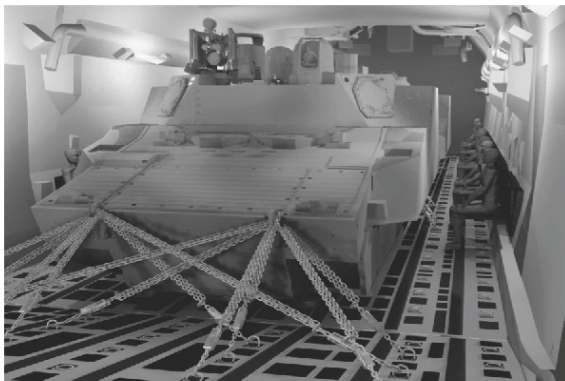


Figure 6. Image from detailed virtual test loading of FCS in C-17 aircraft.

Thus, similar tools and data for heavy lift concept visualization can also be applied to

conducted in great detail for a large number of FCS manned ground vehicle C-17 load combinations from 2008-2009. Although the FCS program was cancelled, the value of insight gained and the expedience of the analysis and feedback is fully established. virtual aircraft test loadings for evaluating the compatibility of large cargo items (military vehicles in particular) with existing and future airlift platforms. The accuracy and certainty of the results depends upon, and grows with, the availability, quality and fidelity of relevant engineering data. This allows heavy lift concept developers as well as vehicle/ cargo developers to make informed decisions and adjustments to the design of a platform without the need and cost of physical prototyping and testing.

4.0 CONCLUSION

High resolution 3D visualizations have been used effectively for both marketing purposes as well as concept development facilitation for heavy airlift concepts. Detailed, accurate 3D models of heavy lift aircraft and large military vehicles are also useful for compatibility analysis and guiding the design process. In summary, the application of common computer aided design and engineering visualization tools along with knowledge, skill and patience can be very fruitful in the development and evaluation of future heavy-airlift concepts and the associated large military vehicle payloads.

5.0 REFERENCES

- [1] Cassidy, J. (1985). *C-130E/H/J/J-30 TEA Pamphlet 70-1 Transportability for Better Deployability*. Retrieved from <http://contracting.tacom.army.mil/>
- [2] United States (2007). *Defense Science Board Task Force on Future Need for VTOL/STOL Aircraft*. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Retrieved from <http://www.acq.osd.mil/dsb/reports/ADA473069.pdf>
- [3] The Boeing Company, Global Mobility Systems (2008). *C-17 Capability Assessment*. Powerpoint File Retrieved from <http://www.boeing.com/defense-space/military/c17/docs>.
- [4] Global Security Organization (n.d.). *JHL Concept Design and Analysis*. Retrieved from <http://www.globalsecurity.org/military/systems/aircraft/jhl-cda.htm>
- [5] Chesnulovitch, D. (2010). *Joint Modular Intermodal Distribution System (JMIDS) Overview [briefing slides]*. Retrieved from <http://www.tatrc.org/ports/medLog/docs/JMIDS-BFG-MEDLOG-Demo-june2010.pdf>

1.6 GPR Imaging for Deeply Buried Objects: A Comparative Study Based on FDTD Models and Field Experiments

GPR Imaging for Deeply Buried Objects: A Comparative Study Based on FDTD Models and Field Experiments

Roger Tilley

University of California, Santa Cruz, CA

rtvax@soe.ucsc.edu

Farid Dowla

Lawrence Livermore National Laboratory

dowla1@llnl.gov

Faranak Nekoogar

Lawrence Livermore National Laboratory

nekoogar1@llnl.gov

Hamid Sadjadpour

University of California, Santa Cruz, CA

hamid@soe.ucsc.edu

ABSTRACT – Conventional use of Ground Penetrating Radar (GPR) is hampered by variations in background environmental conditions, such as water content in soil, resulting in poor repeatability of results over long periods of time when the radar pulse characteristics are kept the same. Target objects types might include voids, tunnels, unexploded ordinance, etc. The long-term objective of this work is to develop methods that would extend the use of GPR under various environmental and soil conditions provided an optimal set of radar parameters (such as frequency, bandwidth, and sensor configuration) are adaptively employed based on the ground conditions. Towards that objective, developing Finite Difference Time Domain (FDTD) GPR models, verified by experimental results, would allow us to develop analytical and experimental techniques to control radar parameters to obtain consistent GPR images with changing ground conditions. Reported here is an attempt at developing 2D and 3D FDTD models of buried targets verified by two different radar systems capable of operating over different soil conditions. Experimental radar data employed were from a custom designed high-frequency (200 MHz) multi-static sensor platform capable of producing 3-D images, and longer wavelength (25 MHz) COTS radar (Pulse EKKO 100) capable of producing 2-D images. Our results indicate different types of radar can produce consistent images.

1.0 INTRODUCTION

Ground Penetrating Radar (GPR) systems have been used to identify voids, pipes, unexploded ordinance and other objects in different soil types. Interpretation of experimental GPR data is often difficult when hampered by varying environmental conditions such as increased or decreased water content in soil. Modeling and simulation techniques can be used to determine the effects of weather and other ground characteristics on the GPR data. Controlled experiments with known ground truth information provides a basis for developing accurate models and a means to develop test beds to study changes in experimental results by varying the model

soil parameters, transmitting frequency, bandwidth, etc. . The Finite Difference Time domain (FDTD) method [4][5][6] has emerged as the leading method of modeling GPR modeling and simulation. Software platforms of interest are FDTD methods that are user-friendly and can run on PC based platforms. Such software platforms can be used to vary dielectric permittivity (ϵ), electric conductivity (σ), and magnetic permeability (μ), the main variables representing the properties of different soil types. The transmitting pulse can also be a variable. There are many such modeling programs that are made available from leading GPR companies and University research centers. We have used the GprMax program, developed by A. Giannopoulos [1], in this paper. In this

study we also used an experimental test site with known objects at known depths to validate the simulation and modeling results. The site was scanned with two types of radar. The first was a monostatic single transmitter (TX) and a single receiver (RX) design; the second system used was a multi-static design with multiple transmitters and multiple receivers. Each “ground truth scan” was processed to remove surface reflections, the global background trace (“system noise”), compensate for the mean or median attenuation, and enhance data by re-sampling the output data at a higher rate. This result was used to compare it with the simulated data from the GprMax FDTD simulation code. Using the site dimensions, the object dimensions, buried depths of each object and soil type a model was created and the simulation results were then compared to the “ground truth” scan results.

This paper is organized as follows. In section 2.0 we describe the field experiments and the test site. In section 3.0 we describe the construction of the FDTD 2-D and 3-D models and discuss the results of each FDTD analysis. In section 4.0 we compare the model results with scanned ground truth data. In section 5.0 we discuss our conclusions.

2.0 Description of the Experimental Tests and Field Data

Target objects were buried in a remote experimental test site, ‘The Forest Lodge,’ located near Greenville, California of the Northern Sierra about 60 miles (96.56 km) north of Lake Tahoe. The objects chosen were metal (tin) roofing sheets approximately 1.83 m (6 ft) long by 66 cm (26 in) wide by 1.27 mm (0.05 in) thick. There were 8 sheets in total, buried at depths of 0.5 (1.64), 1.0 (3.28), 1.5 (4.92), 2.0 (6.56), 2.75 (9.02), 3.0 (9.84), 3.5 (11.48), 4.0 m (13.12 ft) and roughly 1.83 m (6 ft) between sheets. The surrounding soil was a mixture of clay and sand. Figure 1 depicts the tin sheets before burial.



Figure 1. Target GPR imaging objects, tin roofing sheets, were buried at various depths. The experiments provided ground truth GPR data for this study.

The pulseEKKO 100 radar, built by Sensors and Software Inc., was used to obtain the first set of field data. This radar consists of one TX and one RX operating at 25 MHz, making the antennas approximately 3.68 m (12 ft) long by 11.4 cm (4.5 in) wide by 1.6 cm (0.63 in) thick. The antennas were mounted on a low profile platform made of PVC tubing with wheels. The TX and RX antennas were spread apart as necessary to avoid signal saturation by the receiver, shown in Fig. 2.



Figure 2. pulseEKKO 100 radar on a wheeled platform in front of test lane containing buried tin sheets at Forest Lodge.

This radar was used in a mode to generate a 2-D scanned image. Each scan started with the deepest buried object first proceeding to the shallowest object. The transmitting antenna encountered each object first followed by the receiving antenna. The TX and RX antennas were approximately 9.14 m (30 ft) apart.

Examples of the scanned images are shown in Fig.3. Figure 3a and 3b depict the raw data and the processed data respectively.

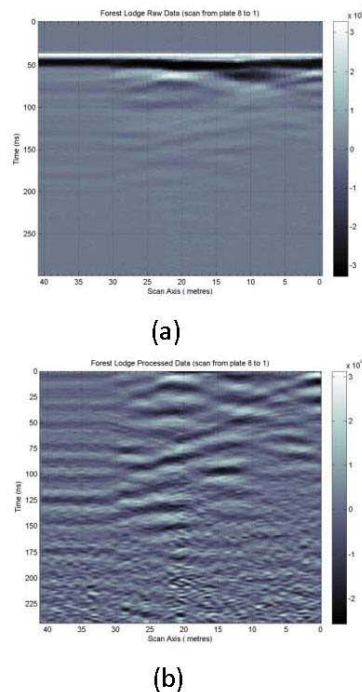


Figure 3. “Ground truth” GPR scan results from the COTS pulseEKKO 100 radar over the Forest Lodge test site at 25 MHz. Raw data is depicted in (a) while processed data is depicted in (b).

The processing used in Fig. 3(b) includes moving the time-zero to remove ground bounce; removing global background to enhance coherent signals and reduce randomly varying signals (noise). Applying an empirical gain function that compensates for the mean or median attenuation observed is part of the image processing. Removing the low frequency component of the data, and lastly, re-sampling the time and scan axes to enhance observed artifacts completes our processing. The axes markings on the pulseEKKO data need some clarification. The axes represent time to target verses distance, scanned over a 30 meter distance. The scan proceeded from 0 meters to 30 m (98.4 ft) beginning with sheet number 8 (the

deepest sheet) to sheet number 1 (the shallowest sheet). The radar was stationary and not part of the testing for scans from 30 (98.4) to 41 m (134.5 ft). For a depth calculation, we multiply time by the assumed velocity in the medium of 0.1 m/ns for the soil type [2], Sand in this case. The existences of voids or objects are clearly shown, some correlate with the tin sheets buried noted in the 2-D model (Fig. 5). The shallowest sheet is approximately at the 22 m (72.2 ft) mark. Other sheets are difficult to distinguish from other artifacts. In the scanned processed results, all known buried objects occur within the first 40 ns using the depth calculation mentioned above. Buried object detail is lacking do to the use of the 25 MHz operating frequency of the radar, but signal response is expected at depths as much as 55 m (180.4 ft) [7].

The 200 MHz multistatic sensor platform was used to scan the Forest Lodge site. This radar consisted of 9 Transmitters and 8 Receivers, fashioned such that each receiver received a signal from 2 adjacent transmitters but, not at the same time forming 2 channels received by one receiver. In this test, the radar provides 16 channels of data that cuts a 2 meter swath over targets of interest to create a 3-D image. The result of the test is shown in Fig. 4, where 5 roofing sheets are clearly present, depicted as stair steps. The depth calculation is the same as in the COTS 2-D pulseEKKO radar case.

3.0 FDTD Models

Modeling of the Forest Lodge site was conducted using GprMax software program developed by A. Giannopoulos described in [1]. GprMax code runs on mainframe computers and Personal Computers (PCs), capable of running compiled C programs like Windows or Linux operating systems. Using techniques described in the GprMax User's manual a 2-D model was constructed as shown in Fig. 5.

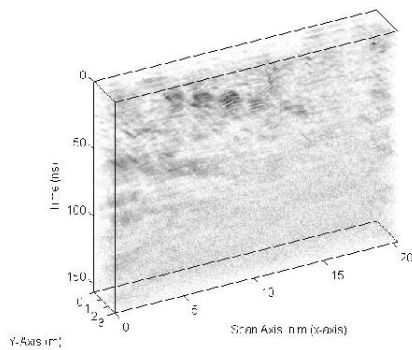


Figure 4. Processed 3-D data scanned by the multistatic radar over the Forest Lodge test site of buried tin sheets of known depth. 5 roofing sheets are visible in a stair step fashion.

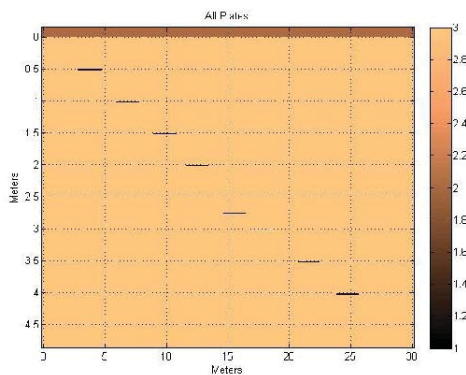


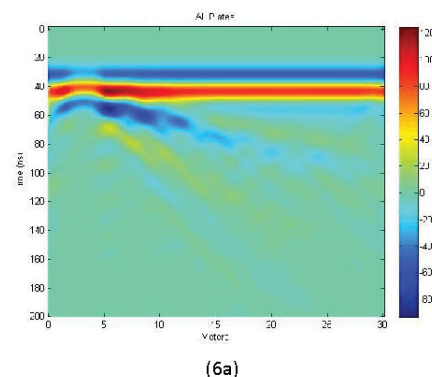
Figure 5. 2-D GprMax model of buried tin roofing sheets, simulating the Forest Lodge test site.

A sample GprMax input consists of defining the media type (sand, clay, water, etc.) and its parameters (dielectric permittivity (ϵ), electric conductivity (σ), and magnetic permeability (μ)). Defining the model boundaries, the spatial step in x and y directions and the time window (the length of time to simulate a GPR trace) are included. Of importance are the dimensions of the buried targets, the excitation waveform, transmitting source(s) and receiver(s) location(s). The final items necessary are the number of analyses to run and the step interval of the source(s)

and/or receiver(s). Units of measure in the model are meters for distance, nanoseconds for time and hertz for frequency.

The 2-D model contains an air space of 0.15 m (0.5 ft) in depth followed by the tin sheets buried at 0.5 (1.64), 1.0 (3.28), 1.5 (4.92), 2.0 (6.56), 2.75 (9.02), 3.0 (9.84), 3.5 (11.48), and 4.0 m (13.12 ft) deep over an area of 30 m (98.4 ft) with spacing between tin sheets of 1.22 (4), 1.04 (3.4), 0.89 (2.92), 1.1 (3.6), 0.33 (1.08), 2.1 (6.89), and 1.3 m (4.27 ft) respectively. Sand was chosen as the medium in which the tin sheets were buried with a relative permittivity (ϵ_r) of 3.0 and an electrical conductivity (σ) of 0.01 mS/m. Two experiments were conducted. The first experiment was conducted at 25 MHz for 1 to 8 tin sheets individually then all tin sheets. The second experiment was run at 900 MHz on all plates to note the response by the model of Fig. 5. Figure 6 depicts the results of the 2-D analysis.

At 25 MHz the object detail is lacking, as expected, though a general trend of the existence of plates is noted. However, the FDTD analysis of individual sheets shown in Fig. 6, demonstrates the effectiveness of this technique in recognizing each sheet but not in detail. Inferred is that the combination of the individual sheet responses results in the trend shown when scanning all sheets at once. At 900 MHz, each plate position is well defined, also as expected [2][4].



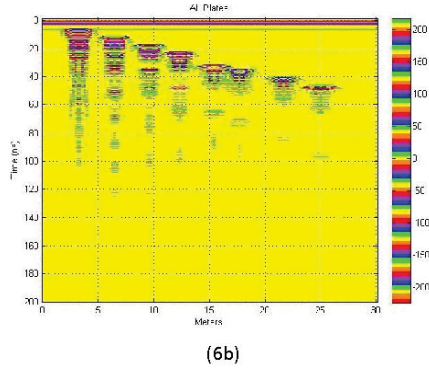


Figure 6. FDTD analysis results of the 2-D model using GprMax at 25 MHz (a), showing minimal detail and 900 MHz (b) showing distinct detail. Unexpected results in (b) are the reflections underneath each tin sheet reflection which has been found to occur in actual scans.

A 3-D model was constructed using the techniques described in the GprMax User's manual. Like the 2-D case, the input file is the same but with added variables to define the third dimension. Included in the 3-D model, as in the 2-D model case, is an air space of 0.15 m (0.5 ft) in depth followed by the tin sheets buried at 0.5 (1.64), 1.0 (3.28), 1.5 (4.92), 2.0 (6.56), 2.75 (9.02), 3.0 (9.84), 3.5 (11.48), and 4.0 m (13.12 ft) deep over an area of 30 m (98.4 ft) with spacing between the tin sheets of 1.22 (4), 1.04 (3.4), 0.89 (2.92), 1.1 (3.6), 0.33 (1.08), 2.1 (6.89), and 1.3 m (4.27 ft) respectively. Each tin sheet is 66 cm (26 in) wide. A value of 2.5 m (8.2 ft) in width was added to the model, completing the third dimensional variable. In the model a matrix of 9 TXs with 8 RXs was constructed without defining which TX and RX communicates at any one time. The completed 3-D model is shown in Fig. 7.

The resulting FDTD analysis of the 3-D model is shown in Fig. 9 depicted at each of the 8 RX antennas. All 8 metal sheets are shown at each of 8 RXs, but some show the reflection from each sheet a little more clearly. An analysis was run at 900MHz for comparison with a more pronounced illumination of the targets as a result.

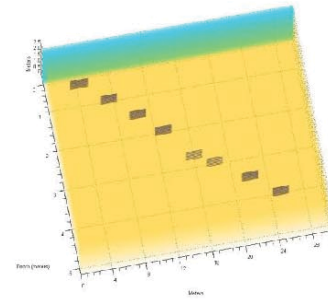


Figure 7. GprMax 3-D model of the Forest Lodge site of buried objects. This model was used to study FDTD response experiments conducted for this study.

4.0 DISCUSSION

Comparing the response of the model at 25 MHz (Fig. 6a) to the actual data (Fig. 3b) taken has proven to be challenging. In general the model and the actual data response is not a good fit. However, the model and actual data have items in common such as the downward sloping trend in the model also noted in the actual data. The downward sloping trend begins at the 22 m (72.18 ft) mark of the actual data, proceeding towards the 15 m (49.21 ft) mark. Another common item is the secondary reflections appearing beneath each tin sheet in the model results as well as in the actual data. Many other artifacts appear in the actual data which were unexpected. We are reminded by the literature that at 25 MHz one can image up to 55 m (180.4 ft) in depth with reduced fidelity [7], which explains the voids at depths well below where our site preparation efforts disturbed the soil. At this time we are without information as to what exists at levels below 4 meters at Forest Lodge. In reviewing the data from the actual 3-D scan at 200 MHz (Fig. 4) we are encouraged that our 2-D response is plausible, because of the clear delineation of targets in the actual 3-D data.

A comparison of the 3-D model (Fig. 7) with the actual data (Fig. 4) at 200 MHz denotes

a better fit. The model locates more tin sheets at lower depths but the actual and modeled data track well. A difference in parameters of the medium easily explains the dissimilarity. A geological survey was not conducted of the Forest Lodge site to determine the specific composition of the medium, we assumed Sand.

5.0 CONCLUSIONS

Ground Penetrating Radar can identify objects buried in different soil types. What we have attempted to show here is that modeling can identify simulated buried objects. One key that is prevalent in the modeling process is that the more accurate the model the better the end result corresponds to the data. A comprehensive knowledge of the area to be modeled strengthens the accuracy of the analysis engine. Our 2-D case had some unexplained outcomes when attempting to compare the real data to the modeled data. Both real and modeled analyses indicated that besides the broad outline of the target there exists a reflection underneath the illuminated object closely associated with just one object at a time. This was an unexpected outcome. For the 2-D case the actual data response was taken over a site that was not well known below 4 meters and appears to have other unexplained voids or objects. The 3-D data bears out that within 4 meters the site content is known. The 3-D analysis shows the buried object more clearly without other artifacts over 4 meters. Part of the difference between the 2-D and 3-D analyses is that one used 25 MHz (2-D case) while the other used 200 MHz (3-D case) center frequencies. Inherent in using 25 MHz is the fact that one will have much less resolution of objects but one will see to a greater depth. Using 200 MHz increases the level of resolution greatly but to a shallower depth. Both models demonstrate that. For the 3-D case, modeling the multistatic sensor platform in its currently defined configuration is challenging. The reason is that the multistatic radar is configured such that 16 channels of data

are returned by a process in which TX and RX communication is controlled. For our model, we took the first step and modeled the 9 TXs and 8 RXs and presented the result. At most we had 8 channels of data where the adjacent TXs added to one RX. The output demonstrated the promise of using modeling as a tool. In general what has been demonstrated is that one can successfully use GprMax to model test sites.

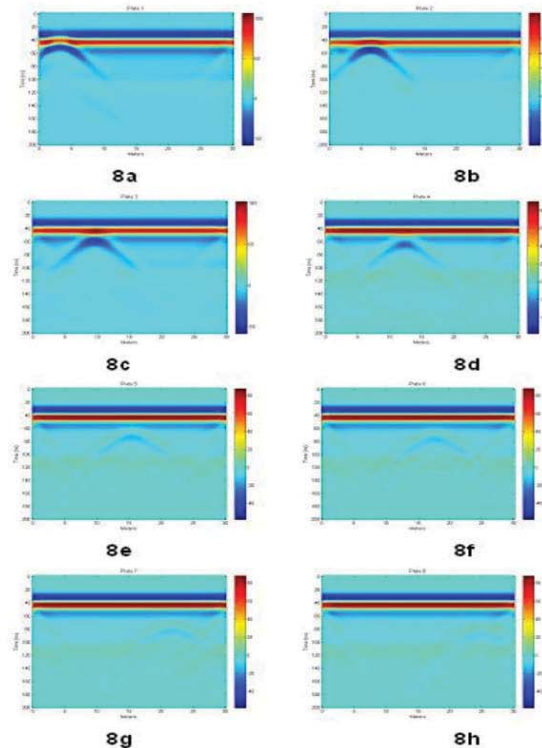


Figure 8. FDTD analysis results from GprMax for 2-D model at 25 MHz, simulating buried roof sheets at Forest Lodge. Scans are of each tin sheet, 8 in all. Tin sheet (a) represents a buried sheet at 0.5 meters, (b) represents 1.0 meter buried sheet. It follows that (c), (d), (e), (f), (g), and (h) represent 1.5, 2.0, 2.75, 3.0, 3.5 and 4.0 meter buried sheets.

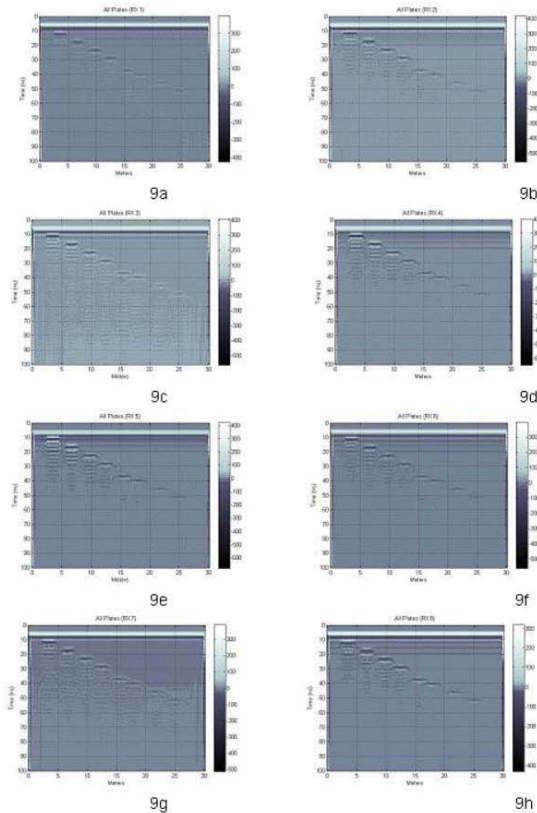


Figure 9. FDTD Analysis results at 200MHz. Plots (a) through (g) depict the software results at each RX modeled in the 3-D GPR model. All 8 of the simulated buried tin sheets are shown. There are unexpected differences from receiver to receiver.

6.0 ACKNOWLEDGMENTS

The Authors thank Dr. Willard H. (Bill) Wattenburg for providing the use of the Forest Lodge facilities; Chico State University Mechatronic Engineering Department for their assistance with field experiments. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-CONF-482511

7.0 REFERENCES

- [1] A. Giannopoulos, "Modelling ground penetrating radar by GprMax," *Construction and Building Materials*, vol. 19, pp. 755-762, Dec 2005.

- [2] A. P. Annan and S. W. Cosway, "Ground Penetrating Radar Survey Design," *Symposium on the Application of Geophysics to Engineering and Environmental Problems*, vol. 5, pp. 329-351, 1992.
- [3] A. Annan and J. Davis, "Ground Penetrating Radar—Coming of Age at Last!!," in *Proceedings of Exploration 97: Fourth Decennial International Conference on Mineral Exploration*, 1997, pp. 515-522.
- [4] A. P. Annan, "Electromagnetic Principles of Ground Penetrating Radar," in *Ground Penetrating Radar Theory and Applications*, M. J. Harry, Ed., ed Amsterdam: Elsevier, 2009, pp. 1-40.
- [5] A. Taflov, "Review of the Formulation and Applications of the Finite-Difference Time-Domain Method for Numerical Modeling of Electromagnetic-Wave Interactions with Arbitrary Structures," *Wave Motion*, vol. 10, pp. 547-582, Dec 1988.
- [6] N. Blindow, *et al.*, "Ground Penetrating Radar," in *Environmental Geology*, Ed: Springer Berlin Heidelberg, 2008, pp. 283-335.
- [7] D. G. Smith and H. M. Jol, "Ground-Penetrating Radar - Antenna Frequencies and Maximum Probable Depths of Penetration in Quaternary Sediments," *Journal of Applied Geophysics*, vol. 33, pp. 93-100, Jan 1995.

1.7 Assessing Tsunami Vulnerabilities of Geographies with Shallow Water Equations

Assessing Tsunami Vulnerabilities of Geographies with Shallow Water Equations

Rifat Aras & Yuzhong Shen
Department of Modeling, Simulation, and Visualization Engineering
Old Dominion University
raras001@odu.edu, yshen@odu.edu

Tsunami preparedness is crucial for saving human lives in case of disasters that involve massive water movement. In this work, we develop a framework for visual assessment of tsunami preparedness of geographies. Shallow water equations (also called Saint Venant equations) are a set of hyperbolic partial differential equations that are derived by depth-integrating the Navier-Stokes equations and provide a great abstraction of water masses that have lower depths compared to their free surface area. Our specific contribution in this study is to use Microsoft's XNA Game Studio to import underwater and shore line geographies, create different tsunami scenarios, and visualize the propagation of the waves and their impact on the shore line geography. Most importantly, we utilized the computational power of graphical processing units (GPUs) as HLSL based shader files and delegated all of the heavy computations to the GPU. Finally, we also conducted a validation study, in which we have tested our model against a controlled shallow water experiment. We believe that such a framework with an easy to use interface that is based on readily available software libraries, which are widely available and easily distributable, would encourage not only researchers, but also educators to showcase ideas.

1.0 INTRODUCTION

Water inundation preparedness is an important component of early response systems for disasters like tsunamis, flooding, and water inundation. Realistic inundation models play an important role in many aspects like finding coastal hazard risks, assessing vulnerable zones, and taking necessary precautions to mitigate damage from inundation.

Shallow water equations are a set of hyperbolic partial differential equations (PDEs) that were introduced 140 years ago [1]. In cases where the horizontal length of the simulation domain is much greater than the vertical length, these equations are obtained by depth integrating the infamous Navier-Stokes equations and widely used to model flows in rivers and coastal areas.

There are many strategies adopted for solving hyperbolic PDEs of shallow water systems, however our particular problem requires a solution to be (1) well-balanced for lake-at-rest problems, in which stationary steady states are preserved, (2) positivity preserving to handle dry and shore areas, and (3) able to support linear friction models. Therefore in this work, we employed a second-order well-balanced

positivity preserving central-upwind scheme for solving the Saint-Venant system [2].

In order for such a model to be useful as a decision support mechanism, it has to comply with certain performance requirements. Faster-than-real time simulation of the model enables decision makers to observe the simulation, try some number of precautionary measures, and choose the optimum one. Unfortunately, it becomes impossible to simulate a sufficiently high resolution model in real time by using only the computational power of a CPU. General purpose computation on graphics processing units (GPGPU) can be described as a paradigm of utilizing high-performance many-core graphics processing units (GPUs) for computation tasks that are normally handled by CPUs. With the transition from fixed to programmable graphics pipeline, software developers gained the ability to use multiple computational cores on a GPU for non-graphics data without the explicit need of managing parallel computation elements such as threads, shared memory, and message passing interfaces [3].

In this work, our main contribution is implementing a system that uses the bathymetry, elevation, and water depth data

of a geography to create the water inundation model. Our proposed system utilizes XNA Game Studio 4.0 [4] to create an interactive experimentation environment to the user. We also harnessed the computational power of GPU in every step of the simulation to be able to run the simulation faster-than-real time to allow decision makers to experiment with certain scenarios and possible outcomes at a fast pace. We have used shaders written in High Level Shader Language (HLSL) in order to delegate heavy computations to the GPU, thus taking the entire load off the CPU that can be used for other tasks.

The remainder of the paper is organized as follows. Section 2 describes the shallow water equations that form the basis of this work and explains the implementation details of the algorithm. Section 3 mentions about the validation methodology of the model and gives details about the simulation experiment we conducted using the presented model. Finally, Section 4 concludes the paper and discusses future work.

2.0 METHODOLOGY

2.1 Shallow Water Equations

The shallow water equations consist of a set of hyperbolic PDEs and can be used to model physical phenomena such as tsunamis, tidal waves, flooding, and other types of water inundation. In vector form, shallow water equations in two dimensions can be written as in Eq. (1).

$$\begin{aligned} \begin{bmatrix} w \\ hu \\ hv \end{bmatrix}_t + \begin{bmatrix} hu \\ \frac{(hu)^2}{w-B} + \frac{g(w-B)^2}{2} \\ hv \\ \frac{huv}{w-B} + \frac{g(w-B)^2}{2} \end{bmatrix}_x \\ + \begin{bmatrix} hv \\ \frac{(hv)^2}{w-B} + \frac{g(w-B)^2}{2} \\ 0 \\ -\alpha hu \\ \frac{-\alpha hv}{1+\beta(w-B)} \end{bmatrix}_y \\ = \begin{bmatrix} 0 \\ -g(w-B)B_x \\ -g(w-B)B_y \\ 0 \\ \frac{-\alpha hu}{1+\beta(w-B)} \\ \frac{-\alpha hv}{1+\beta(w-B)} \end{bmatrix} \quad (1) \end{aligned}$$

The dependent variables in these equations are w (representing surface elevation) and hu and hv (representing horizontal and vertical discharges respectively). B represents the bottom elevation (Figure 1). In a more compact form the equation can be written as

$$Q_t = -F(Q)_x - G(Q)_y + S_b(Q, B) + S_f(Q, B), \quad (2)$$

where Q is the vector of dependent variables, F and G vectors represent fluxes along horizontal and vertical directions, and S_b and S_f represent source terms from the bathymetry slope and friction respectively.

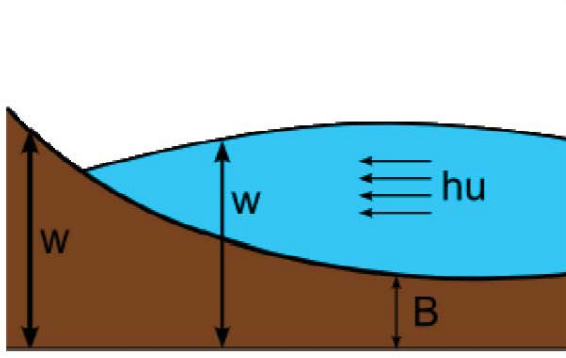


Figure 1. The visual representation of the dependent variables and the bottom topography for the 1D case.

2.2 Central-Upwind Scheme

In order to solve the presented equations, we need a well-balanced positivity preserving scheme. The central-upwind scheme proposed by Kurganov and Petrova [2] is a good candidate, because it is capable of preserving steady states (lake at rest) and also it guarantees the positivity of the fluid depth (near shore areas).

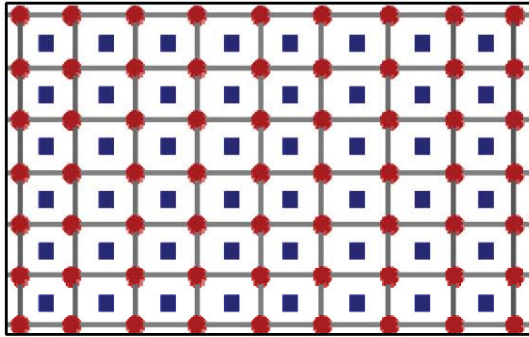


Figure 2. The staggered grid used to store dependent variables (blue squares) and the bottom topography values (red dots).

In this scheme, the dependent variables w , hu , hv , and the bathymetry data B are stored in a staggered grid (Figure 2). The bottom topography is represented as a continuous piecewise bilinear approximation. Likewise, the dependent variables at integration points are also reconstructed from non-oscillatory bilinear approximation obtained from cell averages (Figure 3).

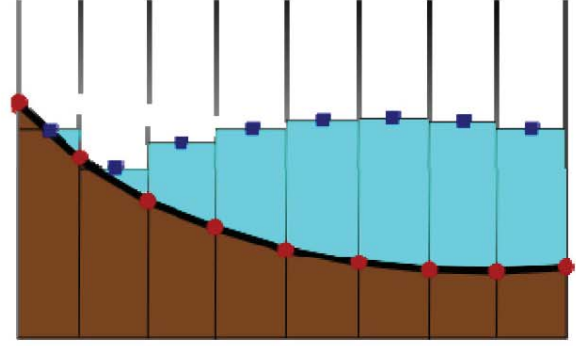


Figure 3. The slopes of the dependent variables are constructed from the cell averages.

To prevent spurious oscillations that would cause instability issues in the simulation, the slopes of the dependent variables have to be reconstructed carefully. Flux limiters, also known as slope limiters, are mechanisms that are used to prevent oscillations, thus making the system solution total variation diminishing. In this work, we used a generalized minmod flux limiter that was originally used by Kurganov and Petrova [2]. The slopes of the dependent variables in one dimension are therefore found by the function,

$$Q_x = \text{MINMOD}(\theta f, c, \theta b), \quad (3)$$

where f , c , and b are forward $(\frac{Q^{i+1}-Q^i}{\Delta x})$, central $(\frac{Q^{i+1}-Q^{i-1}}{2\Delta x})$, and backward $(\frac{Q^i-Q^{i-1}}{\Delta x})$ slope approximations and θ is a parameter that controls how dissipative is the limiter (Figure 4). The MINMOD function is defined as

$$\text{MINMOD}(a, b, c) = \begin{cases} \min(a, b, c), & \{a, b, c\} > 0 \\ \max(a, b, c), & \{a, b, c\} < 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

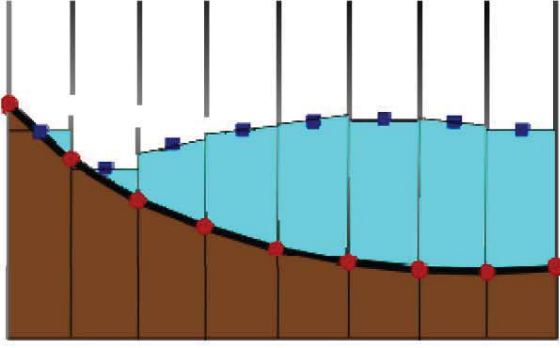


Figure 4. Slopes of the dependent variables are constructed by using a flux limiter.

After slopes are constructed for the dependent variables, one obvious problem with this construction is that we can obtain negative values for the depth component (Figure 4). As the eigen value of the Jacobian of the shallow water equations being

$$u \pm \sqrt{gh}, \quad (5)$$

having negative values for the depth component h will break down the simulation. In order to prevent this from happening, we need to correct the water elevation slopes of the cells that have negative values for the depth component. This is simply performed by moving the problematic negative value up to the topography and recalculating the slope of the cell by using the modified value and the original average value (Figure 5).

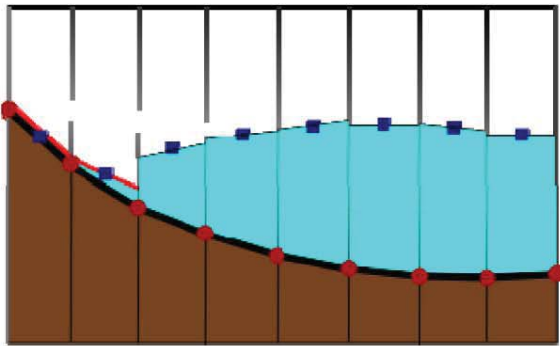


Figure 5. The slopes that result in negative depth values are corrected (red slopes).

After slopes are corrected, the flux values for the cell interfaces are calculated

according to the central-upwind scheme. For the cell interface between cells i and $i+1$, the two neighboring values are obtained by using the corresponding slopes, and these values are used to compute the flux function (Figure 6) (see [2] for details).

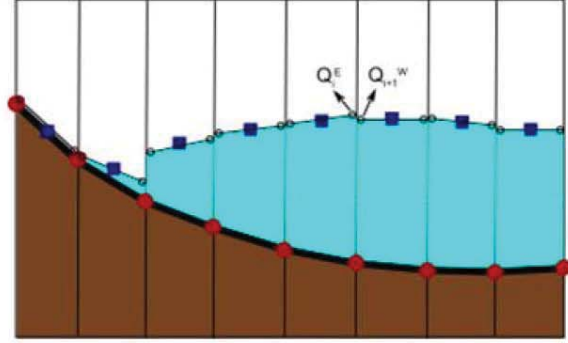


Figure 6. For each cell interface, the two neighboring values are obtained from the slopes to compute the flux value.

Finally, the computed flux values and source terms are plugged into Eq. (2) and time discretization is performed by using the following second order Runge-Kutta ODE solver,

$$\begin{aligned} Q_{ij}^* &= Q_{ij}^n + \Delta t R(Q^n)_{ij} \\ Q_{ij}^{n+1} &= \frac{1}{2} Q_{ij}^n + \frac{1}{2} [Q_{ij}^* + \Delta t R(Q^*)_{ij}], \end{aligned} \quad (6)$$

where $R(Q)$ is the right hand side of the Eq. (2).

2.3 Algorithmic Break Down and Implementation

The central-upwind scheme presented in the previous section can be realized as an algorithm that consists of several passes. A single step of the simulation is broken down into two Runge-Kutta substeps. In each of these substeps, the slopes of the dependent variables are reconstructed, flux computations are performed, and the time integration is computed.

The dependent variables and intermediate values are stored on graphical processing unit as buffers. In XNA programming, these buffers are represented as `RenderTarget2D` objects.

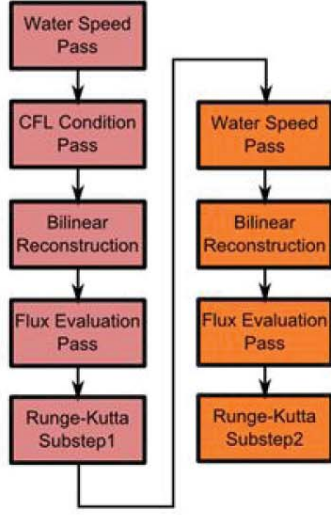


Figure 7. The flow chart of the algorithm.

2.3.1 Water Speed Pass

Water speeds in horizontal and vertical directions are obtained by dividing the dependent variables hu and hv by the water height. In shallow water simulations, fluid speeds are required to determine how large simulation time steps can be and to compute numerical flux values. As we obtain speed values by $\frac{hu}{h}$ and $\frac{hv}{h}$, one apparent problem is having very small height values. Due to floating point precision issues, when we have very small height values, we naturally end up with very high and erroneous speed quantities. This is definitely a problem that would cause very small simulation time steps and can be corrected by desingularizing the computation of speed values for small height values. In this work, we adopted the approach of Kurganov and Petrova [2] and used the following formula to compute speed values at shoal zones:

$$u^* = \frac{\sqrt{2}h(hu)}{\sqrt{h^4 + \max(h^4, \varepsilon)}}, \quad (7)$$

where ε is a small number chosen according to the machine floating point precision. After correcting the speeds for small heights, we need to update the corresponding hu and hv values by $h * u^*$

and $h * v^*$ in order to keep the consistency of the simulation variables.

2.3.2 CFL Condition Pass

The Courant-Friedrichs-Lewy (CFL) condition [5] is a necessary condition for convergence when solving hyperbolic PDEs. For shallow water equations, this condition translates to the condition of simulation time step having to be small enough so that a wave does not pass a whole grid cell during that time step.

In order to find the global fluid velocity value for the entire simulation domain, we need a reduction operation for the water speed values that were obtained from the previous pass. Given a buffer that contains the speed values and whose size is $n \times n$, we can implement the reduction operation as a progressive down-scale operation, in which the new value of the pixel is set to the maximum of the set of pixels that this pixel is sampled from (Figure 8).

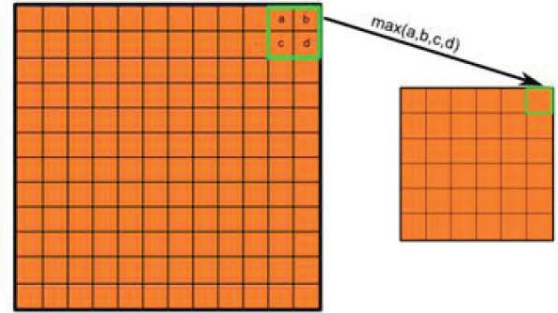


Figure 8. The down-scale operation is performed progressively until we obtain a resultant buffer that is small enough to be processed by the CPU.

2.3.3 Bilinear Reconstruction Pass

The vertical and horizontal slopes of the dependent variables w , hu , and hv are reconstructed by using a branchless implementation of the generalized minmod flux limiter [6] (Figure 9).

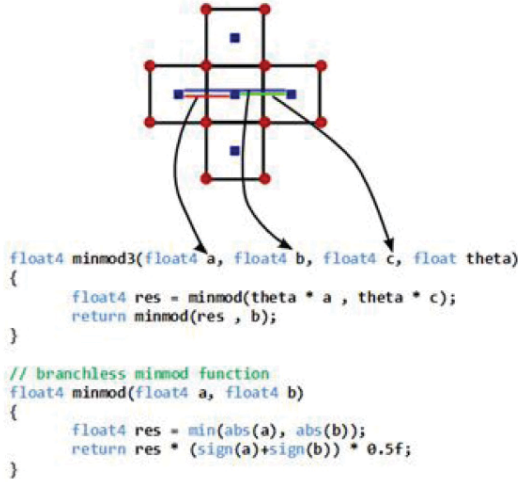


Figure 9. Implementation of the MINMOD limiter in HLSL.

As stated before in section 2.2, special care should be taken after the slopes are reconstructed as these slopes may end up with negative depth values. To prevent this from happening, the midpoints at the cell interfaces are tested against the bathymetry information, if the slope is found to be causing negative depth value, it is corrected.

2.3.4 Flux Evaluation Pass

After the slopes are determined for the grid cells, these are used to approximate the values of the dependent variables at cell interfaces. According to the central-upwind scheme, we need two values for each cell interface that are approximated from the slopes of the two neighboring cells that share the interface.

The horizontal (F) and vertical (G) flux values are computed separately according to

$$F(Q) = \left(hu, \frac{(hu)^2}{w-B} + \frac{1}{2}g(w-B)^2, \frac{(hu)(hv)}{w-B} \right)^T$$

$$G(Q) = \left(hv, \frac{(hu)(hv)}{w-B}, \frac{(hv)^2}{w-B} + \frac{1}{2}g(w-B)^2 \right)^T \quad (8)$$

For implementation details of the central-upwind scheme, refer to the original work of Kurganov and Petrova [2].

2.3.5 Runge-Kutta Substep

In this pass, the flux values computed at the previous flux evaluation pass are used to advance the simulation. The source terms that are caused by bottom topography (bed slope) and friction are also computed and included in this pass. In this work we are using a second order Runge-Kutta solver, in which the simulation is advanced in two separate substeps. In the first substep, the simulation is advanced one time step to obtain the intermediate values for the dependent variables (Eq. (6)). These intermediate values are then used as inputs to the second series of passes to complete the time integration (orange shaded steps in Figure 7).

3.0 VALIDATION & EXPERIMENT

Like any other modeling and simulation study, validity of a model plays an important role in its credibility in decision making processes. In this work, we tried to reproduce the results of a real-life experiment that was conducted in a controlled environment.

In the work of Synolakis et al. [7], the authors emphasize the necessity of verification and validation of the numerical models used in emergency planning. They present a set of analytical, laboratory, and field benchmark tests that can be used to validate numerical water inundation models. In our study, we included their conical island experiment and recreated their experiment setup in our virtual laboratory as accurately as possible (Figure 10).

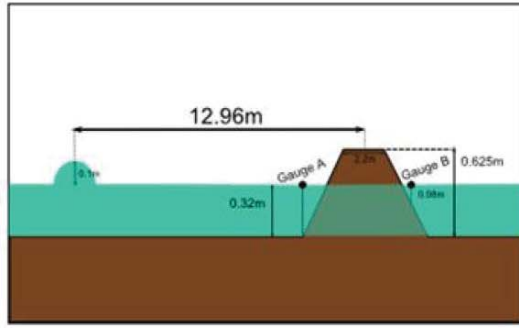


Figure 10. The recreated conical island experiment. Because of the lack of data about wave characteristic, we approximated the wave as a sine wave.

The data from the physical experiment and our virtual experiment are collected and plotted for easy comparison of the two data. Although, there was not enough data about the solitary wave other than its height, we were able to obtain very close results, capturing the essence of the physical experiment (Figure 11).

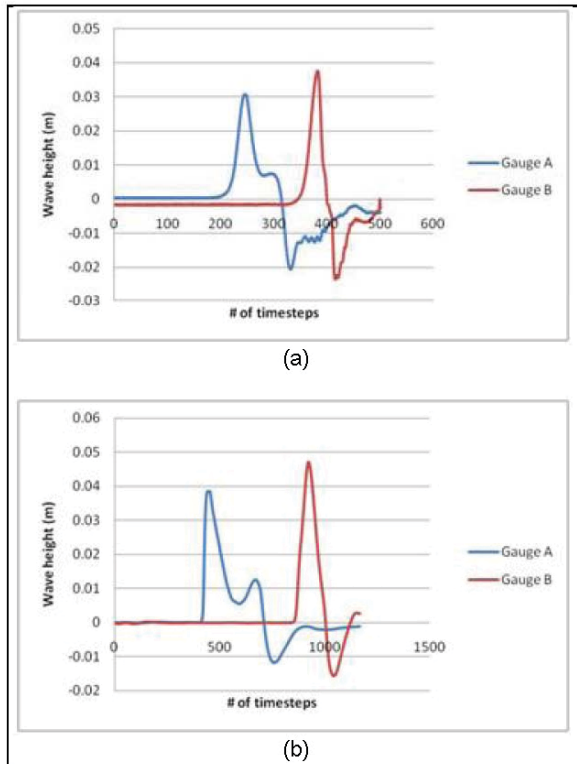


Figure 11. The results of the (a) physical experiment and (b) our simulation.

For showcasing our study, we have chosen the Virginia Beach area, particularly the area around the Virginia Beach Convention Center. The bathymetry and surface elevation data are obtained from National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center[8]. The data, which has a cell size of 10 meters, for the whole Virginia Beach region was preprocessed and converted to comma separated value (CSV) format to be fed to our simulation (Figure 12 and Figure 13).

4.0 CONCLUSION

Recent disasters show that water inundation readiness and emergency planning processes are vital necessities to prevent aftermath. Credible water inundation models are the important elements for emergency planning processes. Decision makers need valid, interactive, and fast models in order to respond to situations in a timely fashion.

In this work, we presented a water inundation model that is based on shallow water equations. We used a well-balanced positivity preserving scheme to solve the set of hyperbolic PDEs. By harnessing the computation power of GPUs, we were able to run a high resolution simulation faster-than-real time on a consumer grade laptop PC. To establish the credibility of the presented model, we recreated a controlled experiment and compared the physical results to our virtual results.

The current model supports static bathymetry / elevation data that is organized as a regular grid. One feature we want to add to our application is the ability to alter the terrain before running the simulation. This way, the decision makers would be able to see the possible effects of certain actions such as adding barriers to parts of the terrain.

5.0 REFERENCES

- [1] A. J. C. d. Saint-Venant, "Théorie du mouvement non-permanent des eaux, avec application aux crues

- des rivières et à l'introduction des vagues dans leur lit," *C. R. Acad. Sci. Paris*, pp. 147-154, 1871.
- [2] A. Kurganov and G. Petrova, "A Second-Order Well-Balanced Positivity Preserving Central-Upwind Scheme for the Saint-Venant System," *Communications in Mathematical Sciences*, vol. 5, pp. 133-160, 2007.
- [3] NVIDIA, "NVIDIA CUDA C Programming Guide," 2010.
- [4] Microsoft. (06.28.2011). *XNA Game Studio 4.0*. Available: <http://msdn.microsoft.com/en-us/library/bb200104.aspx>
- [5] R. Courant, K. Friedrichs, and H. Lewy, "On the Partial Difference Equations of Mathematical Physics," *IBM Journal of Research and Development*, vol. 11, pp. 215-234, 1967.
- [6] T. Hagen, M. Henriksen, J. Hjelmervik, and K. A. Lie, "How to solve systems of conservation laws numerically using the graphics processor as a high-performance computational engine," *Geometric Modelling, Numerical Simulation, and Optimization*, pp. 211-264, 2007.
- [7] C. Synolakis, E. Bernard, V. Titov, U. Kânolu, and F. González, "Validation and verification of tsunami numerical models," *Tsunami Science Four Years after the 2004 Indian Ocean Tsunami*, pp. 2197-2228, 2009.
- [8] L.A. Taylor, B. W. Eakins, K. S. Carignan, R. R. Warnken, T. Sazonova, D. C. Schoolcraft, and G. F. Sharman, "DIGITAL ELEVATION MODEL OF VIRGINIA BEACH, VIRGINIA : PROCEDURES, DATA SOURCES AND ANALYSIS," U. S. D. o. Commerce, Ed., ed. Boulder: National Geophysical Data Center Marine Geology and Geophysics Division, 2008.

6.0 APPENDIX

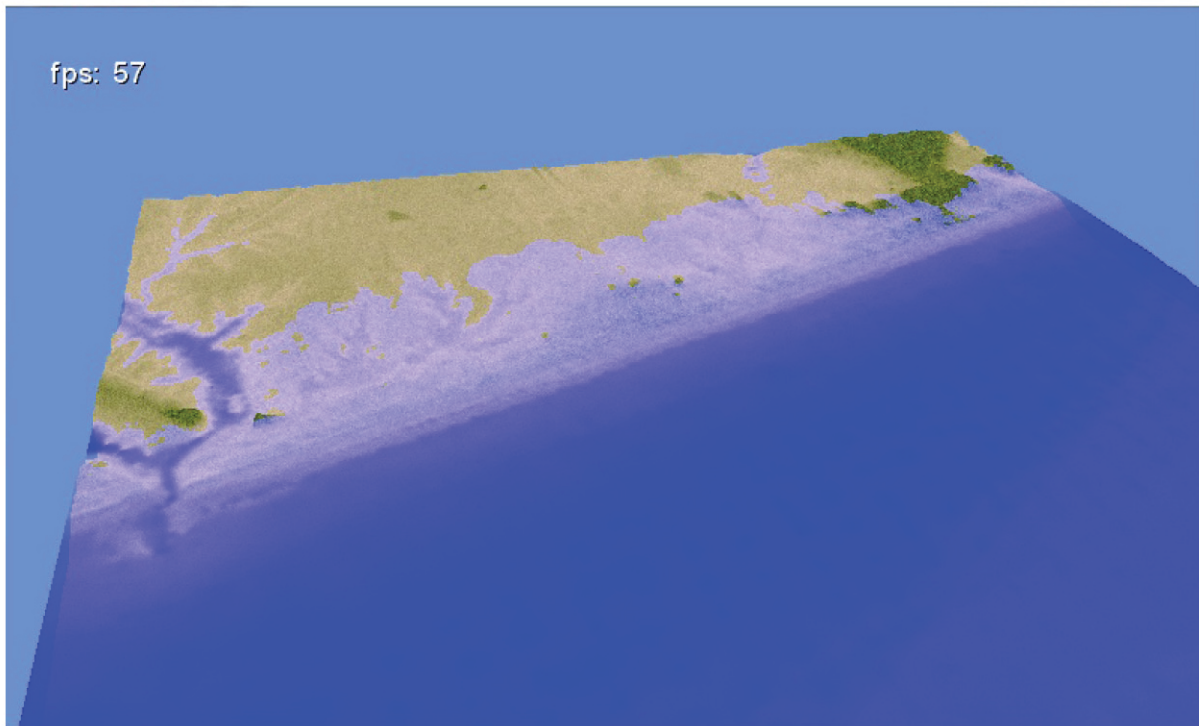


Figure 12. Visualization of the Virginia Beach area simulation as waves hit the shore line.

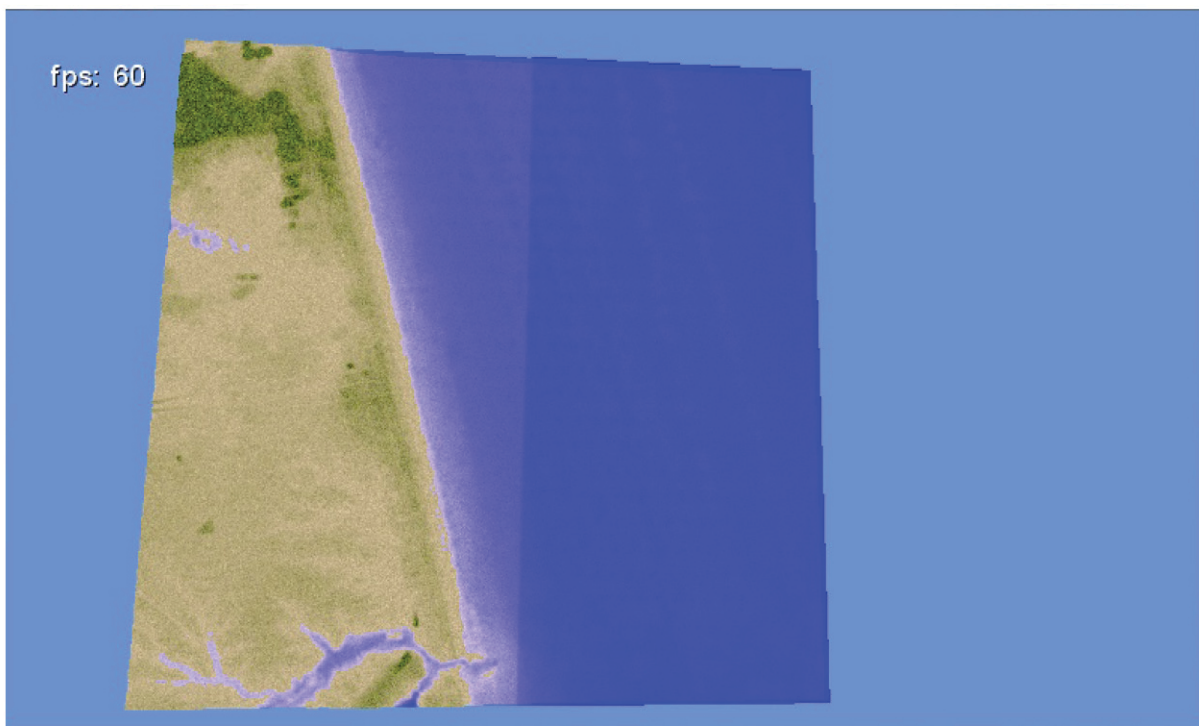


Figure 13. Visualization of Virginia Beach area simulation as waves approach the shore line.



Figure 14: The side view of the wave that approaches Virginia Beach shore line.

1.8 Command Center Training Tool (C²T²)

Command Center Training Tool (C²T²)

Phillip Jones

MYMIC LLC

phil.jones@mymic.net

Nick Drucker

MYMIC LLC

nick.drucker@mymic.net

Reejo Mathew, Ph.D., CMSP

MYMIC LLC

reejo.mathew@mymic.net

Laura Stanton

MYMIC LLC

laura.stanton@mymic.net

Ed Merkle

Virginia Port Authority (VPA)

emerkle@portofvirginia.com

Abstract. This abstract presents the training approach taken to create a management-centered, experiential learning solution for the Virginia Port Authority's Port Command Center. The resultant tool, called the Command Center Training Tool (C²T²), follows a holistic approach integrated across the training management cycle and within a single environment. The approach allows a single training manager to progress from training design through execution and AAR. The approach starts with modeling the training organization, identifying the organizational elements and their individual and collective performance requirements, including organizational-specific performance scoring ontologies. Next, the developer specifies conditions, the problems, and constructs that compose exercises and drive experiential learning. These conditions are defined by incidents, which denote a single, multi-media datum, and scenarios, which are stories told by incidents. To these layered, modular components, previously developed meta-data is attached, including associated performance requirements. The components are then stored in a searchable library. An event developer can create a training event by searching the library based on meta-data and then selecting and loading the resultant modular pieces. This loading process brings into the training event all the previously associated task and teamwork material as well as AAR preparation materials. The approach includes tools within an integrated management environment that places these materials at the fingertips of the event facilitator such that, in real time, the facilitator can track training audience performance and resultantly modify the training event. The approach also supports the concentrated knowledge management requirements for rapid preparation of an extensive AAR. This approach supports the integrated training cycle and allows a management-based perspective and advanced tools, through which a complex, thorough training event can be developed.

1. INTRODUCTION

Our research and development team has developed the Command Center Training Tool (C²T²) to assist the Virginia Port Authority (VPA) in developing and executing training for their dispatchers at their Port Command Center (PCC). The C²T² platform is a training management tool that allows VPA training managers to execute training events for port dispatchers, providing both initial and recurring/refreshers

training. C²T² provides leaders and training managers the means to rapidly create highly complex training exercises, execute those exercises with a minimum of training overhead, and conduct a comprehensive After Action Review (AAR). During execution of the training event, C²T² allows trainers to, in real time, monitor the performance of the training audience, adjust the exercise to meet training objectives, and take advantage of training opportunities. The C²T² approach supports real time

training on any command center system that would otherwise be constrained by time and resources.

C²T² consists of four organizational features; training event creation, training event execution, after action review (AAR), and an administrative feature that allows training managers to better organize and track training events¹.

2. BACKGROUND

C²T² development is based on a Phase II Small Business Innovative Research (SBIR) effort for the Air Force Research Lab's Human Effectiveness Directorate, Warfighter Readiness Research Division. The Phase II product, the Collaborative Management Environment for Team Training (CoMETT), is the core of C²T². Lessons learned during the development and fielding of the CoMETT prototype have been incorporated in C²T² design and development, with specific modifications to meet VPA requirements. The CoMETT system is designed as a platform that supports small teams and teams of small teams. Prototype CoMETT consists of three components. The primary component is the Collaborative Information and Training Environment (CITE), which includes the ability to create, execute, and review a training event. Supporting the CITE is an Advanced Knowledge Base (AKB), which serves as a reference source for CITE users. In the future, CoMETT will also include a Training Management & Collaboration System (TMCS), which will allow a "Community of Training" for the organization that will help identify performance requirements and capabilities and recommend appropriate training interventions to close any gaps between the two.

Due to VPA's unique training requirements, certain components of

CoMETT are either removed or modified. The most prominent change is the modification of the system to function as a tool for training management and not as a tool to facilitate distributed training. While C²T² does not require a distributed capability, the system still supports collaborative interaction through a built-in chat feature. Within C²T², the chat feature is designed to allow training managers and observer/controllers to communicate remotely. However, the system is capable of supporting distributed training, if required by VPA.

Dispatchers at the Port Command Center (PCC) at VPA use Situator™, developed by NICE Systems, to monitor security at all VPA terminals. Situator gathers and correlates information from multiple and diverse systems across VPA and coordinates the response actions that dispatchers are required to perform. C²T² provides VPA with a training interface to conduct training exercises for dispatchers while they are using Situator. C²T² includes advancements to the CoMETT system to facilitate integration with Situator thereby providing a tool specific to VPA's unique requirements.

3. COMMAND CENTER TRAINING TOOL (C²T²)

3.1. Administrative Component

C²T² possess an organizational modeling component that allows VPA to re-create their training structure to include ontological markers of performance. These markers include lists of organizational objectives which trainees must meet, user lists with permission levels to create exercises, execute and monitor exercises, participate as a role player, and master level controls to interface with all aspects of C²T².

The organizational features of C²T² allow training managers to create lists of objectives that can be used to label exercises and scenarios. This includes a component to color code scenarios to make

¹ For the purposes of this paper, the authors use C²T² terminology. C²T² terminology differs from CoMETT's as it is designed to meet the language VPA uses for training.

them readily identifiable during the exercise execution, e.g. red colored scenarios represent fires. Through this organizational modeling component, the manager can create meta-date categories to tag exercises and scenarios, making future searches for required components easier. The labeling capabilities of C²T² also allow the training manger to organize data better for use during assessment components of C²T² training. If the training manager needs to conduct training on a specific requirement designated by the Coast Guard (or any other safety and security governing body), he/she can search for exercises and scenarios tagged with this requirement and either execute the exercise immediately, or make modifications to existing exercises and scenarios and then launch training.

3.2. Exercise Creation

We define four terms for developing training content and executing training within C²T²: incidents, scenarios, exercises, and training events. An incident is an individual trigger, occurrence, or message that acts as a stimulus for the training

audience. Examples include a fire alarm, a door ajar alarm, reported smoke, bomb threat, etc. A scenario is a time-ordered collection of incidents which, when combined, satisfy one or more training objectives. Incidents can be described in a variety of formats, such as documents, emails, audio, video, phone calls, etc. The simplest possible scenario contains a single incident (e.g. a fire alarm). An exercise is a series of one or more scenarios whose totality represents the training exercise. An exercise is an instrument to train for, assess, practice, and improve performance in prevention, protection, response, and recovery capabilities in a risk-free environment. An exercise includes scenarios to respond to and resolve assessed risks. A single scenario may occur more than once in an exercise. A training event is a specific instance of exercise execution in which specific trainees are presented with a selected exercise occurring at a particular time.

Complex exercise scripting is made easy within C²T² with an echeloned scripting process as shown in Figure 1.

Command Center Training Tool (C2T2)

Logout Admin Editor Events Reports Documents User Manual About C2T2

Exercise Library

- MARSEC Exercise
 - Fire Scenario
 - Smell of Smoke
 - Fire Notification
 - Fire Chief Arrival
 - Fire Notification
 - fire out
 - Fire Notification
 - Medical Scenario
 - Medical Exercise
- Scenario Library
- Incident Library

Fire Scenario Incident

Name:

Type:

Owner:

Originator:

Destination:

Time:

Description:

Unknown caller notifies dispatcher that he smelled smoke outside near the dumpster by the construction site outside the rear door of the main building.

make proper notifications, open checklist, annotate daily logs as required.

ExpectedAction:

☐ key incident ☒ Requires Role Player

Figure 1. C²T² Scenarios and Incidents

Exercises are built by combining and sequencing any number of C²T² scenarios. The exercise development process is designed to flow in both directions so that a developer may start with generating scenarios by creating incidents within a timed sequence. They may also begin by creating a set of incidents and building subsequent scenarios and an exercise. A scenario can be a single sequence of incidents, or include branching between different sequences based upon preset decision points. A scenario starts at time "0" and incidents are scheduled by the number of minutes into the scenario, i.e. +2 minutes or +13 minutes. Scenarios also are associated with specific threats to VPA, specific supported performance requirements or organizational tasks, key words, and other meta-data. A scenario, once developed, is loaded to a scenario library, with appropriate security measures.

When the developer desires to create an exercise, he/she can search for scenarios based upon descriptive information. Thus, a developer might search for all scenarios supporting a specific objective/training task. C²T² provides a list of matching scenarios, from which the developer can select and load into his/her exercise. When the scenario is loaded into the exercise timeline, it is assigned a specific start time and all of the incident times are adjusted accordingly as shown in Figure 2. For example, if a scenario is assigned a start time of 35 minutes, then incidents at +2 are scheduled at 37 minutes. The developer can review chosen scenarios, adjust them as desired, or write new scenarios. Training event development is completed by preparing an accompanying simulation and any associated documents that are accessible from inside the C²T² environment.

3.3. Exercise Execution

For effective training, it is important that training facilitators be able to monitor the efforts of the training audience in order to

assess whether the training audience is reaching desired training objectives. With C²T², a training observer now has the added flexibility of communicating a trainee's progress to the training manager using a variety of methods to include a computer, an iPad, or an iPhone. Training facilitators should then be able to adjust the training exercise to make sure that objectives are met and the training benefit is maximized.

C²T² supports training managers and facilitators with a highly configurable interface/dashboard for monitoring and managing the training experience. The prototype CoMETT supported monitoring through the ability to open multiple "gauges" to monitor trainee performance as shown in Figure 3. C²T² allows the training manager to monitor trainee performance by observing a live feed of trainee performance with the Situator software. C²T² then allows the training manager to make notes of trainee performance using both a bookmark feature and scoring sheets currently used to evaluate training. C²T² provides a scrolling event timeline that shows upcoming incidents and scenarios. Training managers can interact with this timeline to reschedule incidents and scenarios, remove or add incidents or scenarios, and pause training if required.

In addition to these "active" tools for trainee performance monitoring C²T² also records trainee performance from the Situator tool itself. Trainee performance, as recorded by Situator, consists of checklists that a dispatcher must complete for each event type. C²T² receives the checklists opened during an event and relays them to the training manager during the AAR to help assess how well the trainee meet the requirements.

Based on guidance from VPA training managers, our team is exploring additional tools or gauges that a training manager may require to accurately monitor and record trainee behavior.

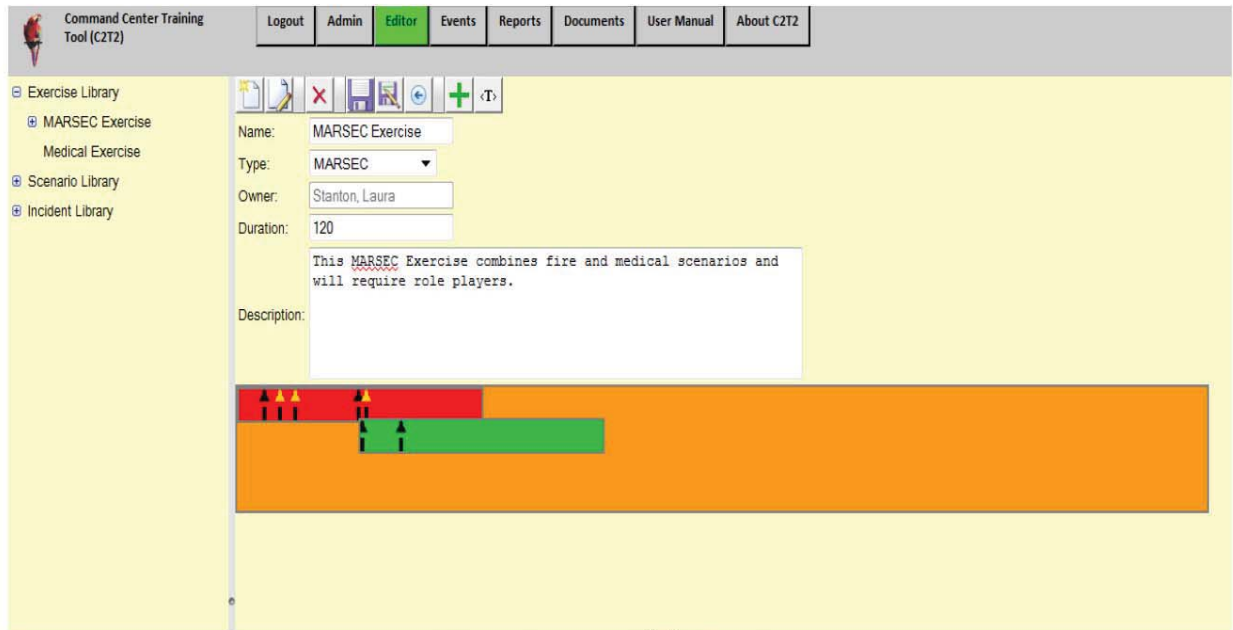


Figure 2. C²T² Exercise

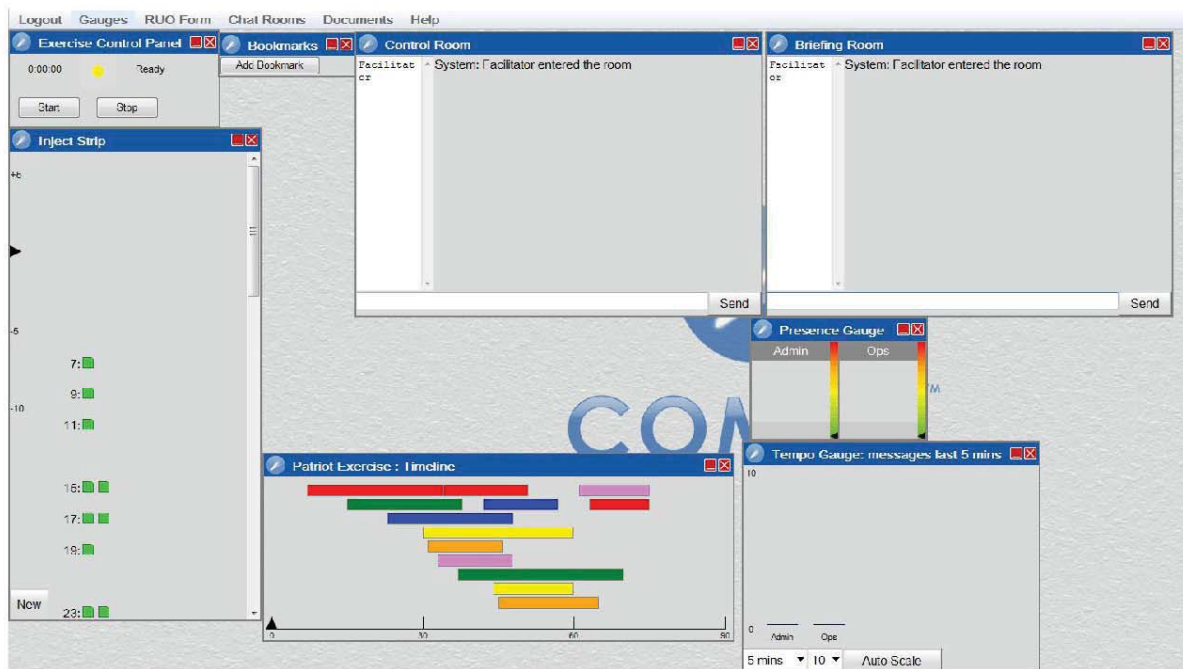


Figure 3. CoMETT Gauges

3.4. After Action Review (AAR)

C²T² includes the ability to conduct an immediate AAR. Following a training event, the training manager can print, or using his computer monitor display: the training event timeline, as it was executed, any training

performance record sheets the manager completed, and the checklists completed by the trainee. VPA requires that trainees cannot access C²T² and, therefore, the AAR occurs either at the manager's desk or at the trainee workstation. During the AAR,

the manager can use their bookmarks, performance assessment sheets, checklists and event timeline to go through the event and talk to the trainee about specific performance.

4. CONCLUSION

The C²T² system is currently in the final stages of development for VPA. The success of the system and the underlying ideas present a new and unique way to manage training events. The ability to rapidly create training exercises, monitor and adjust them on the fly, and track trainee performance for an immediate AAR produce a tool that is both robust and highly suited for training individuals who are required to make decisions in high stress environments. Additionally, the ability to include a collaborative component within C²T² enables future versions of the tool to support VPA or other users in developing training for larger audiences beyond dispatchers, i.e. the ability to create a training event that tests dispatchers, police responders, and other crisis managers within VPA.

5. ACKNOWLEDGEMENTS

This work is funded under the Port Security Grant Program (PSGP) administered by FEMA, with Zel Technologies as the Prime Contractor and MYMIC LLC as a sub-contractor under sub-contract number 1313-MYMIC-10.

1.9 Disaster Response Modeling through Discrete-Event Simulation

Disaster Response Modeling through Discrete-Event Simulation

Jeffrey Wang & Graham Gilmer

Booz Allen Hamilton, Inc.

wang_jeffrey@bah.com gilmer_graham@bah.com

Abstract. Organizations today are required to plan against a rapidly changing, high-cost environment. This is especially true for first responders to disasters and other incidents, where critical decisions must be made in a timely manner to save lives and resources. Discrete-event simulations enable organizations to make better decisions by visualizing complex processes and the impact of proposed changes before they are implemented. A discrete-event simulation using Simio® software has been developed to effectively analyze and quantify the imagery capabilities of domestic aviation resources conducting relief missions. This approach has helped synthesize large amounts of data to better visualize process flows, manage resources, and pinpoint capability gaps and shortfalls in disaster response scenarios. Simulation outputs and results have supported decision makers in the understanding of high risk locations, key resource placement, and the effectiveness of proposed improvements.

1.0 Introduction

1.1 Motivation

Disaster response organizations are faced with the challenges in planning against a rapidly changing, high-cost environment. Incidents can strike with or without warning, and critical decisions must be made in a timely manner to save lives and resources. Essential to making such key decisions is the timeliness of information gathered during the first few hours of a disaster. To minimize decision time and maximize response effectiveness, organizations should understand their overall mission capability and identify key performance gaps and shortfalls that must be addressed.

Government organizations conduct Incident Awareness and Assessment (IAA) missions to provide timely information to first responders. These missions involve the tasking, collection, processing, exploitation, and dissemination of aerial imagery. Images of key disaster zones provide invaluable information to decision makers during a disaster response operation [1].

However, current IAA missions are conducted in an unplanned, ad-hoc manner [2]. In addition, performance requirements have not been adequately quantified and tracked. The current government performance standards have been deemed ineffective and inconsistent across various platforms [3].

Having the ability to consistently measure and track mission performance through time will provide organizations with a comprehensive understanding of current mission capability and identify areas for improvement. The challenge that faces these organizations is devising a flexible and repeatable methodology to quantify performance.

1.2 Approach

We present an analytical approach to quantify IAA disaster response performance and simulate the progression of IAA operations during a disaster response through the use of dynamic modeling.

The disaster response quantification framework and discrete-event simulation presented here is not limited to IAA operations. We see further application towards our approach in search and rescue missions, transportation logistics, and other time-based operational systems that involve a number of moving parts.

2.0 Methodology

The disaster response quantification framework involves researching the disasters in focus, identifying inputs and key relationships, developing performance metrics, designing and running simulations, and analyzing results.

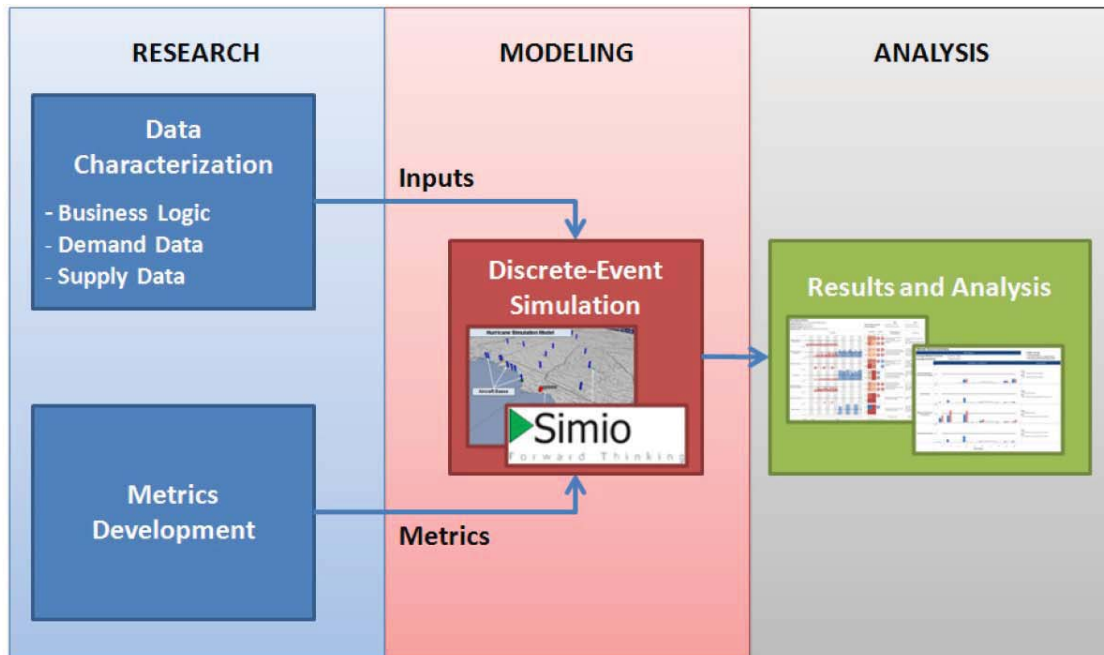


Figure 1. Disaster Response Quantification Framework

2.1 Data Characterization

In order to accurately model and track the performance of current IAA operations, business logic and input data was researched to provide the blueprint for modeling.

The data involved in an IAA Disaster response situation can be separated into two categories, supply and demand. Demand data consists of the overall need for IAA missions to be conducted (i.e. - what the first responders require in order to be successful). Supply data consists of the overall resources that can be drawn upon to satisfy the demand for IAA in a disaster (i.e. - what equipment/platforms can be used to provide IAA).

2.1.1 Business Logic

The disasters in focus were determined to be the following six major types:

- Earthquake
- Hurricane
- Severe Storm/Flooding
- Tornado
- Wildfire
- Winter Storm

To accurately understand the business processes that occur in an IAA disaster response, interviews with various stakeholders and subject matter experts were conducted. These stakeholders and subject matter experts consist of government response forces and first responders that provide IAA support during a disaster incident.

Data collected include:

- Anecdotal data of typical IAA missions
- Major tasks required in a disaster
- Timing during the incident for each required task.

2.1.2 Demand characterization

The demand characterization stage involved researching historical disaster statistics to accurately gauge the magnitude and geographical behaviors of disasters.

Utilizing historical FEMA data, the basic statistics gathered for each disaster incident are as follows:

- Disaster Magnitude

- Counties Affected
- Total Population Affected
- Population Density
- Seasonality

From these elements, the requirements for IAA were determined for each disaster incident across all six disaster types.

2.1.3 Supply characterization

The supply characterization work stream involved research into current IAA aircraft and sensors that perform each IAA task identified in the demand characterization phase.

An extensive database of IAA aircraft and sensor packages was compiled, data included:

- Technical specifications
- Location
- Inventory

In addition, each platform's availability was determined based on their respective agencies, location, and quantity deployed. Availability determines how fast a specific platform could be deployed when a disaster strikes.

2.2 Performance Metrics

Performance metrics were developed to more accurately gauge the effectiveness of the IAA response. These metrics compare how well the platforms and sensors perform when deployed in a disaster scenario.

2.2.1 Penetration

Penetration is the geographic measure of sensor coverage. Penetration is derived from the coverage capability of an individual sensor flying on a specific airframe. When analyzed across a single incident or single disaster, penetration is the weighted average level of coverage for missions given that an aircraft is tasked and deployed.

2.2.2 Persistence

Persistence is the temporal measure of sensor coverage. This metric is two-fold – it

first determines the frequency in which the enterprise is able to provide resources in an incident, if at all. This value allows easy identification of gaps in which aircraft are not available at all. The second step examines, among events where resources are available, the percent of time these assets were devoted to conducting missions rather than flying in-transit, refueling, or awaiting crew rest restrictions. In mathematical terms, persistence is the number of hours resources were conducting missions over the total number of hours of incident response.

2.2.3 Precision

Precision is the qualitative measure of sensor output. Precision is based on whether the coverage satisfied two conditions, optimality and resolution. The first condition identifies whether the sensor is optimally suited to conduct the task based on the operational considerations identified for each task. The second condition is whether the coverage satisfies the minimum National Imagery Interpretability Rating Scale (NIIRS) rating for the required task.

The combination of all three metrics (persistence, penetration, and precision) form Optimal Coverage Level, which is the amount of area to be covered using the optimal sensor at the desired resolution.

3.0 Disaster Response Model – Discrete-Event Simulation

The disaster response modeling framework utilizes discrete-event simulation as its analytical backbone. Discrete-event simulation synthesizes all data elements researched and provides the execution of the business processes involved in an IAA disaster response. The simulations are based on models built using Simio® [4] [5] software developed by Simio LLC.

3.1 Process Model

The disaster response model takes in the supply and demand inputs, simulates the disaster response, and computes the

performance metrics for each historical disaster scenario.

The process model first considers the location, magnitude, and type of disaster and generates the raw mission demand required over the entire disaster period. The demand is parsed out by mission type and represents the total land area that is required to be imaged by the aircraft over the disaster incident.

The model then considers the number of aircraft available, the sensor packages attached, equipment specs, and the geographical location to calculate the overall supply resource pool to draw from. Depending on platform restrictions (range, availability), certain platforms are excluded. Each platform is assigned a number of tasks in which it can conduct and the model calculates the land area coverage rate it can image.

In addition, the model calculates the closest aircraft base for platforms to refuel and redeploy. A buffer is considered such that the closest service station is not within the disaster radius.

3.1.1 Business Logic

Once the variables are calculated, the model follows the following business logic to simulate the disaster response.

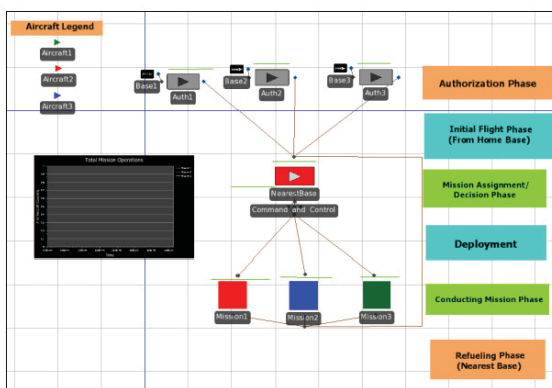


Figure 2. Disaster Response Process Model

Table 1. Model Business Logic

1	Depending on availability, aircraft are released from base and travel to the nearest base
2	Aircraft refuel, and the model assigns a specific task for the aircraft to conduct
3	Aircraft travel to disaster site to conduct mission, collecting various IAA imagery depending on type of task
4	Aircraft runs out of fuel, returns to base
5	Based on current performance metrics and demand, aircraft redeployed to a new task

During the simulation, the model continuously tracks the performance levels of each task required in the disaster. The tasks required may change depending on the current time (as certain missions are needed later in a response). The model will adjust the aircraft deployment logic based on these changes dynamically.

3.2 Graphical View

The Simio modeling software has a 3-D rendering capability, allowing the disaster response model to be displayed in a three dimensional format. In addition, objects in the process model can be animated to visually represent more closely to the real-world counterparts. In the graphic below, the model represents a disaster scenario that is occurring in Florida.

The graphical view provides a method for troubleshooting as well as for presentation purposes to provide a better understanding of the overall simulation logic.

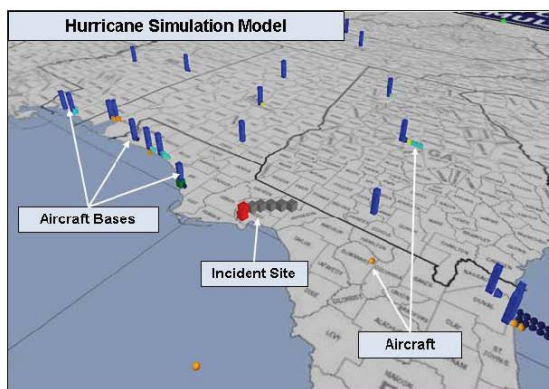


Figure 3. Disaster Response Graphical Model

3.3 Experiments

The IAA disaster response model can rapidly generate a large number scenario runs simultaneously using Simio's experiment feature. This allows the model to run all the historical disaster scenarios researched. Over the past 12 years, there have been over 500 documented federal disaster declarations. The IAA disaster response model simulated all of the scenarios and calculated the performance metrics for each. The metrics are aggregated by geographic region and seasonality to provide an enterprise level analysis of the current IAA response capability.

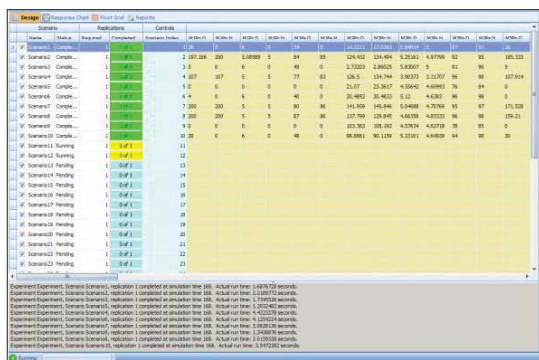


Figure 4. Disaster Response Experiments View

The disaster response model generates .xls outputs of the scenario performance metrics. The persistence, penetration, and

precision values are displayed in dashboard format using MS Excel.

4.1 Single Incident Outputs

Performance metrics for each disaster scenario are displayed using a combination of sand charts, donut graphs, and matrices to highlight missions where there are excess or shortfall in capability.

A shortfall in penetration suggests a lack of available assets to adequately cover the mission demand.

A shortfall in persistence suggests an absence of required assets, and that new equipment or platforms are nonexistent or unavailable in the disaster region.

A shortfall in precision suggests a lack of optimal equipment, and that suboptimal sensors and platforms were deployed due to a lack of appropriate resources.

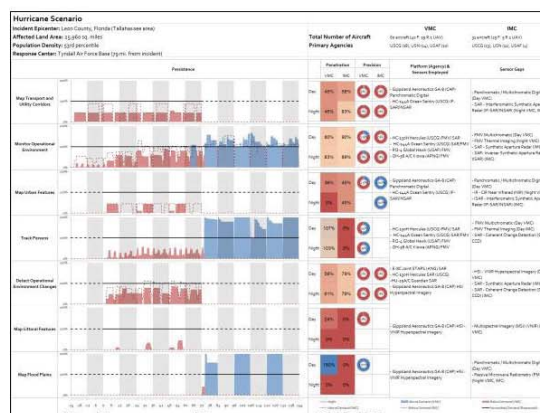


Figure 5. Single-Incident Simulation Dashboard

4.2 Enterprise Level Outputs

In addition to outputting performance metrics for individual scenarios, the disaster response model also generates enterprise-level scenario data by disaster type.

Data from each of the six incident types is aggregated geographically by FEMA region into an enterprise dashboard. Persistence metrics for each incident task are measured

across each region to identify under equipped regions. The metrics are also compared to the frequency of incidents within each region to provide the severity of a capability gap. Smaller gaps bear more weight in a region frequently affected by an incident than a larger gap occurring in a region less affected.

4.0 Conclusion

Dynamic modeling with discrete-event simulations provided a consistent, robust, and repeatable framework for the analysis of IAA disaster response operations. The application assisted the study in understanding the current gaps and shortfalls in current resourcing. Data generated from simulations helped pinpoint geographically the problem areas within the overall enterprise that need additional attention. In addition, dynamic modeling allowed the study to better visualize their current business processes, and also provided a talking point for discussion.

In the future we plan to expand the disaster response models to further applications. Such applications include forest fire resourcing operations, cost analysis, and also disaster imagery Processing, Exploitation, and Dissemination (PED) architecture analysis.

5.0 References

- [1] 601st AOC ISRD/ISRO. "ISR Ops Playbook: IAA and DSCA Ops Disaster Response," June 11, 2008.
- [2] Punaro, Arnold L., William L. III Ball, Les Brownlee, Rhett B. Dawson, Larry K. Eckles, Patricia L. Lewis, Dan McKinnon, et al. "Commission on the National Guard and Reserves - Final Report," January 31, 2008.
- [3] Deptula, Lt Gen David. "Air Force ISR in a Changing World, Changing Paradigms While Optimizing 'Low Density' to Meet 'High Demand'." PowerPoint presented at the 48th AIAA Aerospace Sciences Meeting including the New Horizons Forum and

Aerospace Exposition, Orlando, FL, January 2010.

[4] Pegden, Dennis C. "Intelligent Objects: The Future of Simulation".

[5] Simio LLC. "Introduction to Simio," 2010.

6.0 Acknowledgments

The authors thank Michael Fox and Dr. Dennis Pegden for their subject matter expertise and guidance.

1.10 Modeling a Civil Event Case Study for Consequence Management using the IMPRINT Forces module

Modeling a Civil Event Case Study for Consequence Management using the IMPRINT Forces Module

Marc Gacy, Mala Gosakan, Angela Eckdahl, Jeffrey R Miller
Alion Science and Technology

agacy@alionscience.com, mgosakan@alionscience.com, aeckdahl@alionscience.com, jrmiller@alionscience.com

Abstract A critical challenge in the Consequence Management (CM) domain is the appropriate allocation of necessary and skilled military and civilian personnel and materiel resources in unexpected emergencies. To aid this process we used the Forces module in the Improved Performance Research Integration Tool (IMPRINT). This module enables analysts to enter personnel and equipment capabilities, prioritized schedules and numbers available, along with unexpected emergency requirements in order to assess force response requirements. Using a suspected terrorist threat on a college campus, we developed a test case model which exercised the capabilities of the module, including the scope and scale of operations. The model incorporates data from multiple sources, including daily schedules and frequency of events such as fire calls. Our preliminary results indicate that the model can predict potential decreases in civilian emergency response coverage due to an involved unplanned incident requiring significant portions of police, fire and civil responses teams.

1. BACKGROUND

Consequence Management(CM) is one critical area of non-combat operations where the Army works with civilian and government agencies such as the Federal Emergency Management Agency (FEMA) for sustained periods of time. A challenge in the CM domain is the ability to allocate manpower and resources to deal with complex and dynamic situations while ensuring that personnel with the right skill sets and levels of readiness have been assigned to the jobs at hand. One such software tool that gives CM planners the ability to see the impacts of manpower and resource allocation on operational readiness is IMPRINT.

1.1. IMPRINT

IMPRINT is a simulation and modeling tool that provides means for estimation Manpower, Personnel, and Training (MPT) requirements and to identify constraints for new weapon systems early in the acquisition process. The IMPRINT tool grew out of common U.S. Air Force, Navy, and Army MPT concerns identified in the mid-1970s (Dahl, 1990). It is government owned software and consists of a set of automated aids to assist analysts in conducting human performance analyses (<http://www.arl.army.mil/IMPRINT>). IMPRINT has been available as a government product free of charge since the mid 1990s to the following organizations - U.S. government agencies, U.S. private industry with U.S. government contract, and U.S. colleges and

universities working in Human System Integration. IMPRINT has four modules - Warfighter, Missions, Equipment, and Forces module that provides a powerful and flexible environment in which to develop human performance models, and has unique capabilities for assessing the impact of stressors (e.g., sleep deprivation, protective gear) on performance (Allender, 1999). This is achieved through an embedded simulation engine, based upon the commercial Micro Saint Sharp

(<http://www.alionscience.com/index.cfm?fuseaction=Products.view&productid=35>)

discrete event simulation tool and supplemented by human performance algorithms. In this paper we describe the case study model used to exercise the Forces capabilities of IMPRINT.

2. FORCES APPROACH AND CASE STUDY MODEL

The Forces module in IMPRINT allows the analyst to perform manpower projections for large groups of people by analyzing the successful completion of both planned and unplanned activities. The level of detail is sufficient to be meaningful down to the individual level, but abstract enough that tradeoff analyses are simple to develop and perform.

In order to exercise these capabilities and to create a better understanding of the kinds of data that can be generated from this kind of analysis, we developed a case study model

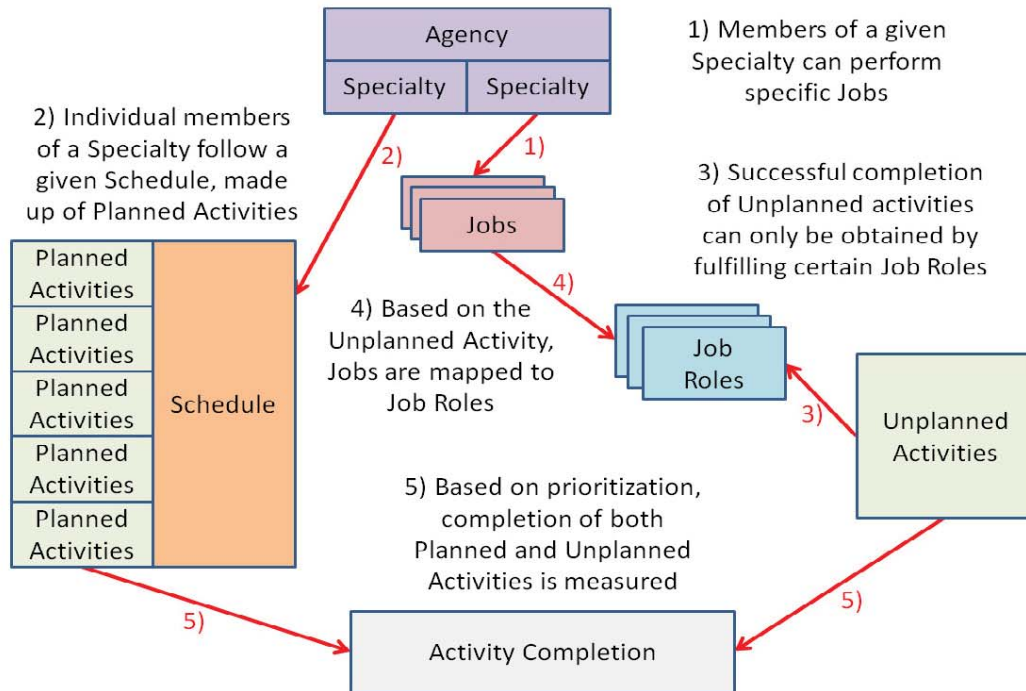


Figure 1: Interplay among Components of the Forces Module of IMPRINT

at a sufficient level of detail and at an appropriate level of organizational structure that represents both regular and extraordinary activities using both local and national agencies.

The case study involves a hypothetical college campus terrorist threat incident and its impact on civil service agencies. In our model, which represents a five day period, the potential bomb threat occurs near the beginning of the third day.

2.1. Scenario Description and Overview

A small city is three days into a typical week with the fire department having responded to several fire calls, when what would normally be a routine civil disturbance call escalates into a situation involving four separate agencies over a 16 hour period. An initial call to deal with a troublesome incident on a campus ends with the student being apprehended by police. During this time a

potentially biohazardous packet of powder was found, resulting in a quarantine of all involved personnel. A Civil Support Team with experience and equipment were notified to respond, entered the area and obtained samples. The Civil Support Team also investigated another related incident-a suspicious package, believed to be a bomb found in an adjacent building. The entire time, a local support team of EMT, Fire and Police personnel were assigned to assist with securing the area and providing support to the Civil Support Team. During this time another fire call came which was met successfully, but with fewer than the normal number of personnel. A mutual aid fire call from a neighboring county came in, but the fire department did not have the personnel to support the call..

2.2. Interplay among Force Components

The interplay among the various components of the Forces module of IMPRINT is shown in

Table 1 Agencies and their Specialties used in the Case Study

Agency	Specialties
Fire Department	<i>Chief, Driver, Ladder Operator, Pump Operator</i>
Police Department	<i>Chief, Detective, Patrol Officer, SWAT Team Leader, SWAT Team member</i>
Emergency Medical Technician	<i>EMT Driver, Transport Care Giver</i>
Dispatcher	<i>Switch Board Operator</i>
NBC Recon Civil Support Team	<i>NBC Recon Team leader, NBC Recon NCO, NBC team member</i>

Figure 1, with the net result being the measurement of activity delays and completion failures. The numbered arrows correspond with the relationships described in the figure.

Thus using the prioritization scheme among planned and unplanned activities and assigning personnel to fulfill specific roles, the ability to successfully complete activities is calculated.

The Forces module also contains components for assessing equipment usage for both planned and unplanned activities, but as they are functionally similar to Specialties, Jobs and Job Roles, we are omitting discussion of them in this paper for brevity and clarity.

2.3. Agency Definition

Five agencies were detailed and used in the case study model. Each agency has a specific number of specialties, indicated above in (Table 1). In this paper we discuss in more details below, one agency, *NBC Recon Civil Support Team*.

2.3.1. NBC Recon Civil Support Team

The *NBC Recon Civil Support Team* specialties are based on discussions with a previous member of the 7th Civil Support Team and an analysis of the 7th Civil Support Team's Standard Operating Procedures (7th CST SOP). The *NBC Recon Team leader* is the officer in charge of the team, being responsible for the development of overall plans and being responsible for all section operations. The *NBC Recon NCO* is responsible for coordinating and reporting

team actions back to the NBC Team Leader. The *NBC team member* carries out the plans of the Team Leader and the NCO. There is a further distinction of duties in the 7th CST SOP of Recon Team Chief, but that is delineated in *Jobs*, below.

2.4. Jobs

Jobs represent specific Job assignments from the given specialties. These are mapped to *Job Roles* described below.

2.5. Schedules

Schedules were developed for each of the agencies based on interviews and information obtained from members of the various organizations. The schedules were developed and the planned activities were pulled from each of the schedules, renamed as necessary to common terms (e.g., "lunch" in one schedule would become the generic "Eat" for all schedules") A sample schedule is shown below (Table 2).

Table 2 Sample Schedule

Planned Activity	Start	End	Duration
Sleep	0.0000	6.0000	6.0000
Personal Time	6.0000	7.0000	1.0000
Physical Training	7.0000	8.0000	1.0000
Hygiene	8.0000	9.0000	1.0000
Classes/Training	9.0000	12.0000	3.0000
Eat	12.0000	13.0000	1.0000
Standard Work	13.0000	17.0000	4.0000
Personal Time	17.0000	22.0000	5.0000
Sleep	22.0000	24.0000	2.0000

2.6. Planned Activities

Planned activities are the various events that are used to populate Schedules.

Planned activities were generated based on the development of the schedules described above. They are not represented in detail; rather they are given a priority which is compared against the priority of Unplanned Events (below) to determine which activity is performed at any given time. In this analysis, due to the acute nature and severity of the situation, no planned activity takes precedence over any unplanned activity. Finally, an activity may be a "Sleep Activity" which is used to determine who may be called to perform a specific duty. Given equal task priorities, a person who has had more sleep will be chosen to perform a task over someone who has not.

A sample list of the Planned and Unplanned Activities sorted by priority (lower the value, higher the priority), with the Sleep Activity indicated is shown below (Table 3).

Table 3 Sample Planned/Unplanned Activities (NBC Team)

Priority	Name	Type
1	3) Quarantine	Unplanned
1	C) NBC Recon offsite training	Unplanned
2	4) NBC Recon Support	Unplanned
2	4) NBC Recon - powder	Unplanned
2	5) NBC Recon - bomb	Unplanned
97	Eat	Planned
104	Hygiene	Planned
105	Sleep	Planned
108	Personal Time	Planned
112	Standard Work	Planned
115	Physical Training	Planned
116	Classes/Training	Planned

2.7. Job Roles

The *Job Roles* are defined as roles that need to be performed for successfully addressing

the Unplanned Activities of the scenario. A specific *Job* may fulfill one of several potential *Job Roles*, and *Job Roles* may be performed by any number of *Jobs*. A brief description of each of the responsibilities of a person performing a given *Job Role* follows (Table 4). A sample of the cross referencing of Jobs and Job Roles as discussed above (Figure 1) is shown below (Table 5).

Table 5 Sample Jobs Cross-Referenced with Job Roles

NBC Jobs	NBC Comms	NBC Decon	NBC Leader	NBC Member
Recon NCO	X	X	X	X
Recon Team	-	X	-	X
Recon TL	X	-	X	-
Team Chief	-	X	X	X

2.8. Unplanned events

There are two sets of unplanned events. Those that relate to the terrorist threat call (numbered in the naming convention adopted in the model) and those that relate to other unplanned activities (lettered in the model).

Unplanned activities may be either fixed duration, or a variable duration that can depend on the number of people available to handle the unplanned activity. In this analysis, the major events are fixed times, since they lead from one directly into the other while the minor unplanned activities (Fire Calls) have a variable duration.

2.8.1. Terrorist Threat Call

The Terrorist threat call has several components which comprise the whole incident, described below.

2.8.2 Initial Terrorist Call

This is the initial Police Response to the disturbance on campus. It is a half hour

Table 4 Sample Job roles and their descriptions used in the case study

Job Role	Description
NBC Communication	Track and communicate with NBC team while in a hazardous area
NBC Decontamination	Decontaminate NBC team that entered the hazardous area.
NBC Leader	Coordinate NBC entry into the hazardous area.
NBC Member	Enter hazard site, or be part of backup team.

unplanned event during which the *Dispatcher*, one *Police Leader* and two *Police Members* are involved. One *Police car (Police Mobility)* is involved. The suspect is found and apprehended by police.

2.8.3 Call Followup

After the police apprehended the suspect, various emergency personnel were called to the scene, including four more police officers (including two *Police Members* and two *Police Sub-leaders*), two *EMTs* and four from the Fire Department members (including two *Fire Members*, one *Fire Leader* and one *Fire sub-leader*). This requires one *Ambulance*, the *Command Center*, another police car (*Police Mobility*) and two fire vehicles (*Fire Mobility*) to bring the various members to the scene. This procedure takes an hour and a half, during which it becomes known that there is a package with a suspicious powder that may be a biohazard threat (e.g., anthrax).

2.8.4 Quarantine

All of the personnel who have come into contact with the suspect in the room and with the powder are quarantined in the building. This is several of the people from the *Call Followup*, but not all. It includes both *EMT* and six of the seven police officers (all but one of the *Police Members*). No vehicles are quarantined. Note that the Quarantine has the highest possible priority in this analysis, as once the personnel are quarantined, they cannot be released for another activity until the quarantine is lifted. The quarantine will last until officially lifted, nine and a half hours later.

2.8.5 NBC Recon – Powder

Once the potential biohazard threat has been identified as an issue requiring additional attention and the proper clearances have been given, the NBC Recon Civil Support Team was notified and responds to the scene. They arrive on the scene three and half hours later. They bring their equipment truck and trailer (*NBC Mobility*), along with eight personnel (including one person for

NBC Communication, one *NBC Leader*, two *NBC Decontamination* and four *NBC Members*).

In addition to *NBC Mobility*, four *NBC Personnel Protection* suits were used (two for entry, two for decontamination), and two sets of *NBC Detection Equipment* (one for entry one for backup) were required. Including transport, arrival, setup, entry and decontamination, this activity takes five hours.

2.8.6 NBC Recon Support

Local service support is required to assist the NBC Recon Civil Support Team with various tasks, including crowd control, emergency services, and providing access to buildings. This involves more coordination with the *Dispatcher*, an additional *EMT*, the *Fire Leader* (if available), the *Police Leader* (if available), one or two *Police sub-Leaders*, one or two members of the Fire Department (*Fire Member*) and four or five members of the Police department (*Police Member*). The SWAT team comprises the bulk of the police force in this activity, so it requires a SWAT van (*SWAT Mobility*), in addition to a police vehicle and fire vehicle (*Police Mobility* and *Fire Mobility* respectively). Finally coordination among all the services requires both the use of the base *Command Center* and the on-location *Mobile Command Center*.

This activity is numbered the same and NBC Recon – powder, since they both start at the same time. They are kept as separate activities so it is easy to modify them for our sensitivity/tradeoff analysis below. The length of this task spans the total time of both NBC Recon tasks, seven hours.

2.8.7 NBC Recon – Bomb

Near the time when the biohazard threat is being wrapped up, another suspicious package (a potential bomb) is found in a nearby building. Even though this activity requires more personnel (using five *NBC Personnel Protection* suits instead of four), the team is already on-site and setup, so this

activity only takes two hours. Other than using the one additional suit, this activity uses the same *Job Roles* and *Asset Features* as 4) *NBC Recon – Powder* above.

2.9. Other Unplanned Activities

There are three other Unplanned Activities that are taking place during the time around the potential terrorist threat. Two are two different kinds of fire calls, and one relates to the NBC Recon crew, listed below.

2.9.1 Local Fire Call

A Local Fire Call occurs when a fire emergency is reported in the general reporting area for the Fire Department. We used an average of one fire call per day, using a Log Normal distribution, with a 10 hour standard deviation. Normally the standard deviation could be larger to allow for greater variation, but we wanted to ensure that a Regular Fire Call would occur during the terrorist threat call. The call requires two *EMT*, one *Fire Leader* (if available), three or four *Fire Members*, one *Police Leader* and one *Police sub-leader* (if available), and one *Police Member*. The vehicles required for a Regular fire Call include one *Ambulance*, one or two *Fire Mobility*, one *Ladder* (if available), one *Water Pump*, one police car (*Police Mobility*) and a *Mobile Command Center*.

Also for the *Local Fire Call*, the time of the call can be reduced by increasing the number of crew members. In this case each crew member reduces the time by 15%, up to a total contribution from four extra crew members of a 48% reduction in time $[1 - (1 - 0.15)^4]$. This represents the increased efficiency and availability of members to communicate,

2.9.2 Mutual Aid Call

A Mutual Aid call occurs when a fire emergency is reported that is outside the general reporting area for the Fire Department. Due to the nature of these call

being outside the normal operating area, the calls typically take longer and require more fire personnel (so a second vehicle is less likely to be sent) but fewer fire support personnel. Two more Fire personnel (*Fire Member*) are required, but one less *EMT* is needed and a *Police Chief* is not needed, as these are supported by the originating locale. Mutual aid calls occur approximately once every eight days depending on the time of year. We placed our mutual aid during the terrorist threat incident to force a consequence in our analysis. Also notice that the priority is lower than a regular fire call, since it is being called from its general reporting area, a *Local Fire Call* would take precedence over a mutual aid call.

2.9.3 NBC Recon Offsite Training

The members of the NBC Recon Civil Support Team occasionally take extensive off-site training which makes them physically unable to respond to a call. For this analysis, three members of the NBC Recon team (*NBC Member*) were unavailable for the duration of the threat call.

3. ANALYSIS

3.1. Overall assessment

This is a fairly short incident with a focused response. The analysis here mostly relates to resource management of personnel and equipment. Other than the failed Mutual Aid Call, the local services and NBC Recon Civil Support team are able to address and handle the situation without much of a disruption of normal activities. The *Local Fire Call* that occurs during the *Call Followup* and immediately before the quarantine is the most affected activity that still succeeds. It uses seven people, one less *Fire Member* and none of the other desired roles. This is close to the minimum required number of six, such that the response time is not significantly reduced as most of the other Local Fire Calls are by the presence of extra personnel.

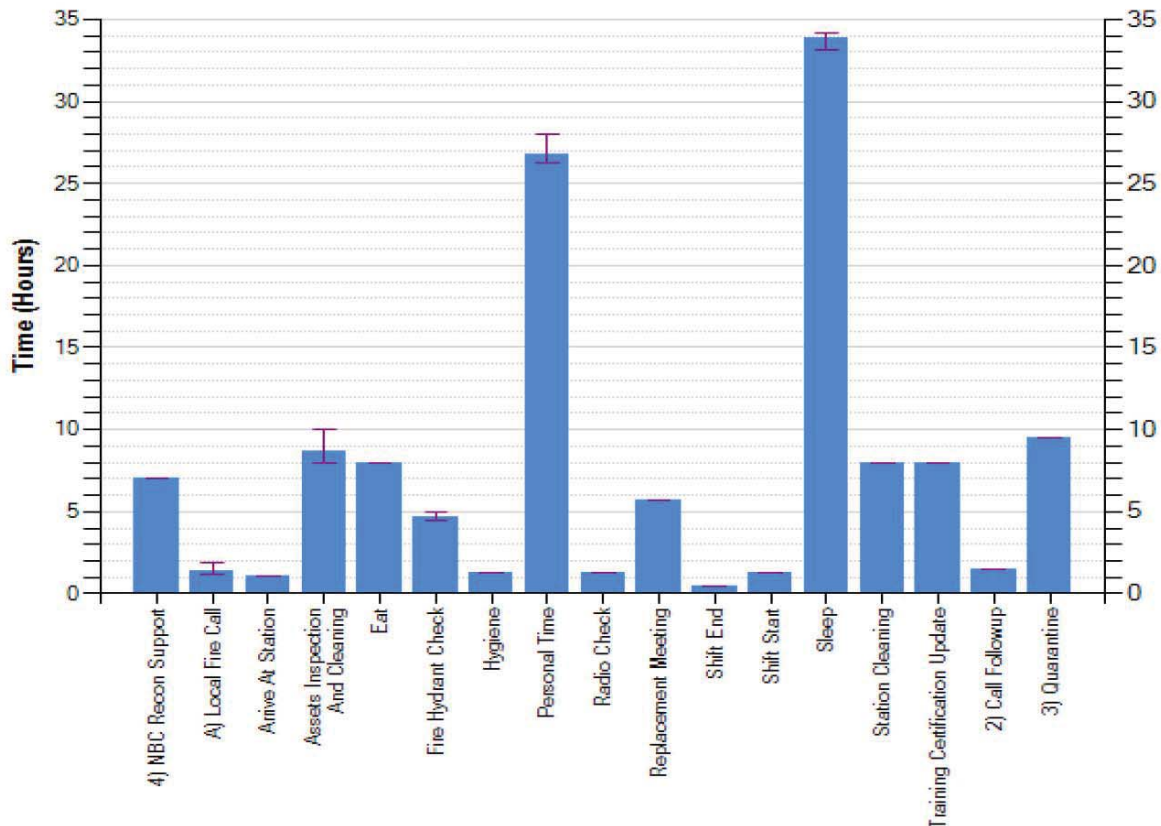


Figure 2. Mean and Std Dev Times for Planned Activities

3.2. Effect of increased crew members

The Local Fire Call activity is the only activity in this analysis that is affected by the number of crew members. There are four instances of Local Fire Calls, which have three different numbers of responders among the four instances.

The first Fire Call is reduced in time by 39% due to the presence of three extra crew members (9 desired – 6 required). The second call is reduced by the full 48% described above, due to having all four extra crew members (9 desired – 6 required). The third call, however only shows a 15% improvement in time due to the fact that there is only one extra crew member. Here, the maximum number of crew members that can affect the time (four) is also equal to the extra desired crew members (four). If the desired number of crew members were to increase to 11 or more, it is important to note that the times would remain unchanged, since the maximum number of crewmembers that can

affect the time is still four. If however, the desired number stayed the same but the maximum number of crewmembers can affect the time were reduced to three, the second fire call would see only a 39% reduction in time, instead of the indicated 48%.

3.3. Time of effects

While many of the graphs and data indicate a small to moderate impact of the threat incident on various duties and other planned activities, there are a few things to point out. First, in the Fire Department Driver Activity chart below, both Personal Time and Asset Inspection and Cleaning show more variation ranging from 26 to 28 hours and 8 to 10 hours respectively. While the change in personal time is not drastic, over time it could become prominent. What is also of concern is that Asset Inspection and Cleaning is reduced. During an extended incident, this activity could suffer to the point where asset failure may be more likely to occur.

Similarly, in the Activity times for Police Dept Officers, patrol time is negatively impacted by the threat incident (Figure 2). Again, while in the short term of this analysis, a reduction in patrol time of a few hours a week may not seem much, increasing the required number of police or extending the time of the incident to much longer periods of time would have consequences on the quality of the police coverage in the area.

3.4. Consequences of terrorist threat call

The terrorist threat activity prevented the Mutual Aid Call (at 80.0 hours), due to the support required for the NBC Recon Activity. This call would likely have been handled by another Fire Department adjacent to the Fire Department which initiated the call.

4. TRADEOFFS

4.1. Extended quarantine

In this scenario, the initial biohazard tests come back positive, and the quarantine is extended to two days. A larger police and fire support team is required and for a longer period of time, due to an increase in the geographic area and to keep a growing population of onlookers from the site.

4.1.1 Changes

The Quarantine is extended to two full days, the NBC Recon Support is extended to a day and a half and expanded to two required personnel and four desired personnel for both Fire and Police Members.

4.1.2 Results

During this period, in addition to the failed mutual support call, two Local Fire Calls failed (89.9 hours and 109.1 hours) due to the extended time and requirements of the NBC Support. The initial threat has ended, but due to the uncertainty of the situation, the quarantine continues well into normal daytime hours. This would generally lead to word being spread of the incident and an increase in onlookers. This results in a larger group being required to keep the situation in control.

The extra fire personnel required means that two Local fire calls are missed: one that occurs almost immediately after the NBC Recon activities are finished and another near the end of the quarantine time. Here the anticipation of the quarantine extending into the day suggest an anticipatory response to bring in more fire personnel.

4.2. More Frequent Local Fire Calls

In this scenario, the terror threat call occurs at a time when Local Fire Calls are more common (summer months).

4.2.1. Changes

The frequency of Local Fire Calls is doubled, changing the Mean Time from one day to twelve hours and a standard deviation from ten hours to five hours.

4.2.2. Results

In addition to the Mutual aid call failing at 80 hours, a Local Fire Call that comes in during the NBC Recon activities (81.9 hours into simulation) fails, as Fire personnel are still supporting the NBC Recon effort. This indicates that when Fire Departments are near capacity (greater frequency of responses), an unplanned event is more likely to have negative consequences. This would indicate that during such times, contingency plans such as alerting adjacent fire districts of the higher likelihood of a mutual aid call or preemptively requesting more available volunteers (if a volunteer fire department) is warranted.

5. CONCLUSIONS

In this paper we demonstrated the uses of the Forces module through the use of a case study involving several distinct agencies, including both civilian and military civil support.

The case demonstrated that the level of detail is specific enough to observe substantive changes in task completion times, impacting overall manpower estimates for both task completion and quality of life (sleep and personal time cycles).

6. ACKNOWLEDGMENTS

The research reported in this document was performed in connection with contract -Prime Award # W911NF-07-2-0062 and Sub-Award # LWI-291-018 - for the Leonard Wood Institute and U.S Army Research Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The authors would like to acknowledge Sgt. Juan Gallego for his contributions to this effort.

7. REFERENCES

Allender L et al., "Evaluation of human performance under diverse conditions via modeling technology. In Improved Performance Research Integration Tool (IMPRINT), User's Guide (Appendix A). Aberdeen Proving Ground, MD: U.S. ARL", (1999)

Dahl (now Archer) et al., "Final Report for Concepts on MPT Estimation (Development of MANPRINT Methods, Report E-17611U, Prepared for US Army Research Institute for the Behavioral and Social Sciences", December 1990.

Gosakan, M., "Enabling CM Response Force Readiness through HSI Tool Enhancements," Final Report for sub awards number LWI-291-018, February 2011.

1.11 Cognitive Systems Modeling and Analysis of Command & Control Systems

Cognitive Systems Modeling and Analysis of Command & Control Systems

LTC Arne Norlander MSc PhD
Swedish Armed Forces
Joint Concept Development and Experimentation Centre
arne.norlander@gmail.com

Abstract. Military operations, counter-terrorism operations and emergency response often oblige operators and commanders to operate within distributed organizations and systems for safe and effective mission accomplishment. Tactical commanders and operators frequently encounter violent threats and critical demands on cognitive capacity and reaction time. In the future they will make decisions in situations where operational and system characteristics are highly dynamic and non-linear, i.e. minor events, decisions or actions may have serious and irreversible consequences for the entire mission. Commanders and other decision makers must manage true real time properties at all levels; individual operators, stand-alone technical systems, higher-order integrated human-machine systems and joint operations forces alike. Coping with these conditions in performance assessment, system development and operational testing is a challenge for both practitioners and researchers. This paper reports on research from which the results led to a breakthrough: An integrated approach to information-centered systems analysis to support future command and control systems research development. This approach integrates several areas of research into a coherent framework, Action Control Theory (ACT). It comprises measurement techniques and methodological advances that facilitate a more accurate and deeper understanding of the operational environment, its agents, actors and effectors, generating new and updated models. This in turn generates theoretical advances. Some good examples of successful approaches are found in the research areas of cognitive systems engineering, systems theory, and psychophysiology, and in the fields of dynamic, distributed decision making and naturalistic decision making.

1.0 INTRODUCTION

According to Ref. [1], forces deployed in future expeditionary Crisis Management Operations (CMO) will need to rely on extraordinary capabilities to operate in austere and distant theatres of operations, on interaction and collaboration within and between different organizational cultures, between people with different backgrounds, education and experience, and on managing and maintaining technological and doctrinal effectiveness and interoperability.

CMO – be they military missions, counter-terrorism or emergency response – comprise complex, laborious and dangerous tasks, requiring resolute and determined teamwork under extreme conditions [2]. Deployed units must also be able to operate independently and with little support while still ensuring operational security and mission efficiency without risking excessive resource depletion.

CMO teams incorporate numerous human and artificial team members, widely scattered across the whole theatre of operations. They can operate autonomously for certain time periods and in

specific areas, but they are mostly forced to coordinate their actions very accurately with one another.

New technology will offer extensive support to every phase of CMO. The enormous potential of prospective high-capacity information processing and real time interaction in distributed, dynamic mission environments is yet to be fully exploited.

Additionally, success in future military and emergency response missions requires highly capable understanding, defined as *the perception and interpretation of a particular situation in order to provide the context, insight and foresight required for effective decision-making* [3], enabling every commander and operator to develop a comprehensive appreciation of the situation and a detailed system insight, leading to safe and efficient mission accomplishment.

Finally, the turbulent environment in which these units operate stresses the need for Organizational Agility (OA), defined as *“the capacity to be infinitely adaptable without having to change... The goal is to keep internal operations at a level of fluidity and flexibility that matches the degree of turmoil*

in external environments, a principle known as requisite variety.” [4]. This requires adaptive and versatile principles and concepts for Command and Control (C2) along with agile high-performance organizational structures, with broad and deep support from strategic, operational and tactical doctrine.

2.0 AN INTEGRATED COGNITIVE SYSTEMS MODELING FRAMEWORK

Highly capable Command and Control (C2) support is needed for omnidirectional, continuous interaction and information exchange between the executive level and the team-on-site. However, as Rochlin [5] and others have observed, the specific skills and properties that systems, managers and operators have to possess in order to yield optimal mission performance in such critical and uncertain situations are not easily identified. Hence, they are difficult to improve. *Action Control Theory (ACT)* [6] is a conceptual framework specifically composed to facilitate modeling and analysis of complex dynamic tactical systems and processes and of their states and state transitions.

The different research areas of ACT have until now developed along separate paths of evolution. Flach & Kuperman [7] concluded that it is essential to develop a unified and proactive approach in research and systems design for future warfare environments. ACT is a composite theoretical structure designed to generate comprehensive and robust models of systems, tasks and missions, supported by advanced experimental and measurement methods, and data analysis techniques. It can support complex, multi-level human-machine systems design in the military, aviation and emergency response domains [8].

2.1 Cognitive Systems Engineering

The area of Cognitive Systems Engineering (CSE) has grown at an increasing pace since the first significant contributions were

published in the 1980s by Rasmussen [9; 10], who introduced the concept of skill-based, rule-based and knowledge-based behavior for modeling different levels of human performance. Hollnagel & Woods [11] made a significant contribution to this field by their definition of a Cognitive System (CS) as a Man-Machine System (MMS) whose behavior is goal-oriented, based on symbol manipulation and uses heuristic knowledge of its surrounding environment for guidance. A CS operates using knowledge about itself and the environment to plan and modify its actions based on that knowledge. According to Hollnagel [12], the definition has been revised over the years in order to comprise new findings in human-machine systems research and to cover a more comprehensive and fundamental set of system properties: what the system achieves, what objectives it serves and what its intentions are. The current CS definition is *a system that can modify its pattern of behavior on the basis of past experience in order to achieve specific anti-entropic ends* [12]. Military forces are often analyzed and modeled as aggregates of Doctrine, Organization, Training, Materiel, Leadership and education, Personnel, Facilities and Interoperability (DOTMLPFI), all of which are elements of the total operational capability. Viewing this capability together with its operational context as a CS facilitates modeling and analysis with significantly greater breadth and depth.

2.2 Complex Dynamic Adaptive Systems, Control Theory and Cybernetics

From the work of Ashby [13], Brehmer [14] and many others it is well known that most complex systems have real-time, dynamic properties; the system output at a given time is not only dependent of the input value at this specific time, but also on earlier input values, and that a good regulator of a system has to implement a model of the system that is to be controlled. Put otherwise, Ashby's law of requisite variety [13] states that the variety of a controller of a

dynamic system has to be equal to or greater than the variety of the system itself. By the term *dynamic system* is meant an object, driven by external input signals $u(t)$ for every t and as a response produces a set of output signals $y(t)$ for every t [6].

CMO possess highly dynamic and nonlinear operational system characteristics, i.e. small actions or decisions may have serious and irreversible consequences for the mission as a whole [15]. Modern literature describes the broader aspects of defense systems and CMO in terms of Complex Adaptive Systems (CAS) [16; 17] in the sense that crisis management forces or emergency response organizations demonstrate CAS properties, and identifies adaptive mechanisms at the levels of adaptive systems, capability development and collective/society, which adjust through learning, evolutionary development and cultural change to fulfill an externally imposed purpose.

The concepts of control theory and CAS can be used as metaphors in research on decision making, especially in multiple-player, dynamic contexts. The mathematical stringency and powerful formalism of control theory makes it possible to describe and analyze systems as diverse as technical, organizational, economic and biological systems. Orhaug's [18] notion that decision making constitutes the regulatory function in command and control processes strongly supports the control theory approach. This notion also supports the fact that the hierarchical command structures of military and emergency response organizations are strongly coupled to both centralized and distributed decision making principles [19]. Annett [20] used control theory to investigate team skills. Four fundamental requirements must be met [21; 22; 12] if control theory is to be used in analysis and synthesis of dynamic systems:

1. There must be a goal (the goal condition).
2. It must be possible to ascertain the state of the system (the observability condition).
3. It must be possible to affect the state of the system (the controllability condition).
4. There must be a model of the system (the model condition).

2.2.1 Controlling Joint Systems and Processes

The combined view of control theory in technical as well in behavioral domains is crucial for success in this research area.

To this could be added two widely used concepts, originally presented by Rasmussen [10] and adapted to the context of modeling and analysis of C2 systems:

Level of abstraction: At what level/s of organizational or functional abstraction are functional C2 requirements satisfied and C2 capabilities implemented?

Level of aggregation: At what level/s of organizational or functional aggregation are functional requirements satisfied and capabilities implemented?

When a function is A) *implemented* at one level of abstraction, B) *represented* at a second level of abstraction and C) *controlled* at a third level of abstraction the requirement for timely and complete information varies accordingly. On the other hand, it is not important whether a function or mission is carried out by an operator or by an automated system under higher-order supervision, the commander still need to be able to handle ill-structured problems and develop an accurate *appreciation* [23] – going beyond what is normally defined as developing *situation awareness* [24].

C2 in CMO is prone to *model error*. It is unavoidable in such dynamic and uncertain situations. The necessary adjustments and updates of the controller's internal system

model can be made by constantly measuring the deviation of the system output from the reference value. The joint cognitive system is unstable without feedback, and thereby feedback will be needed to correct deviations and compensate for the incompleteness and inadequacy of the internal system model. Reason [25] emphasized the importance of balance between feedback (reactive) control and feedforward (proactive) control. This concept is crucial to achieve optimal C2 performance in a tactical mission.

2.3 Decision Making in Complex Systems Control and Mission Command

Brehmer [14] suggested the use of control theory as a framework for research in Distributed Dynamic Decision Making (DDDM). The conventional view of decision making, supported by normative theories, reduces decision making to selecting an appropriate action from a closed, pre-defined action set, and to resolution of conflicts of choice. As a consequence, the analysis of decision tasks focuses on the generation of alternatives and the evaluation of these alternatives as in Multi-Attribute Utility (MAU) analysis [26]. Research in dynamic decision making has instead been based on analysis of several applied scenarios, e.g. military decision making, operator tasks in industrial processes, emergency management, aviation and intensive care which Brehmer [14; 19] summarized as follows:

1. A series of decisions is required to reach the goal.
2. The decisions are mutually dependent.
3. The state of the decision problem changes over time.
4. The decisions have to be made in real time, requiring true dynamic and adaptive capabilities throughout the organization.

2.3.1 Naturalistic Approaches to Decision Making

Zachary & Ryder [27] reviewed decision making research during the last decades and elaborated on the recent major paradigm shift in decision theory. The shift is from analytic, normative decision making procedures described in Kleindorfer et al. [26] to Naturalistic Decision Making (NDM), developed and described by Klein [28; 29], Zsombok & Klein [30] as well as by Klein & Woods [31]. NDM applies to many dynamic and potentially dangerous areas of activity such as military missions, air traffic control, fire fighting, emergency response and medical care [32]. The essentials of this paradigm are condensed below:

1. Human decision making should be studied in its natural context.
2. The underlying task and situation of a problem is critical for successful framing.
3. Actions and decisions are highly interrelated.
4. Experts apply their experience and knowledge non-analytically by identifying and effecting the most appropriate action in an intuitive manner.

2.3.2 Tactical Team Decision Making

According to Cannon-Bowers et al. [33], tactical decision making teams in modern warfare or crisis management are faced with situations characterized by rapidly unfolding events, multiple plausible hypotheses, high information ambiguity, severe time pressure, and serious consequences for errors. There are also cases when geographical separation or other forms of distributed environments in which the teams operate impose additional difficulties. To be able to adapt to these situations, team members must co-ordinate their actions so that they can gather, process, integrate, and communicate information timely and effectively. This is particularly true in CMO where it is difficult to assess performance

with a single correct answer or in situations where several individual decision makers must interact as a team.

Another important aspect is how the actual mission and the mission environment affect team performance. Serfaty & Entin [34] drew the following conclusions concerning the properties and abilities of teams successfully performing tactical, hazardous operations:

1. The team structure adapts to changes in the mission environment.
2. The team maintains open and flexible communication lines. This is important in situations where lower levels in a command hierarchy have access to critical information not available to the higher command levels.
3. Team members are extremely sensitive to the workload and performance of other members in high-tempo situations.

2.4 Psychophysiology

Within joint cognitive systems performing complex, high-risk military and emergency response missions there is a fundamental and profound connection between human operator physiological stress response and discrepancies between expectancies and experiences. Levine and Ursin [35] describe stress response as a warning of an occurring homeostatic imbalance.

This implies that the concept of model error from control theory once again can be applied. The stress response is also mobilizing physiological resources to improve performance, which is regarded as a positive and desirable warning response. The Cognitive Activation Theory of Stress (CATS) describes the phases of the stress response as an alarm occurring within a complex adaptive cognitive system with feedback, feedforward and control loops, no less but no more complicated than any other of the body's self-regulated systems [36]. The time dimension of stress responses must be accounted for very carefully.

2.5 Models of Action Control

The point of departure in our ACT-based systems modeling endeavor was the definition of a Tactical Joint Cognitive System (TJCS) [4] as the system

- To which a mission is assigned.
- To which the operational command of the mission is appointed.
- To which the responsibility for effecting the mission is handed.
- To which the resources needed for performing the mission are allocated.

A TJCS is an aggregate of one or several instances of four principal sub-system classes:

1. Technological Systems, for example vehicles, intelligence acquisition systems, communication systems, sensor systems, life support systems, including the system operators.
2. Command and Control Systems, consisting of an information exchange and command framework, built up by technological systems, rules, procedures and protocols, and directly involved decision makers.
3. Support Systems, comprising staff functions, logistic functions, decision support functions, organizational structures, doctrine, Tactics, Techniques and Procedures (TTPs) and various kinds of service support.
4. Tactical Teams, composed and defined according to Salas et al. [37] as: "Two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have been assigned specific roles or functions to perform, and who have a limited life-span of membership."

The TJCS forms the basis for the Tactical Action Control Model (TACOM). The

principal components of the TACOM are the Mission Environment, the Tactical Joint Cognitive System, the Situation Assessment function, and the Cognitive Action Control function.

The next step is integration of these concepts into a Mission Execution and Control Model (MECOM). The MECOM consists of one or several TACOMs extended with control theoretic components, to handle system disturbances, model error, and to allow an adaptive and balanced mix of feedforward and feedback control.

The last step in the model formation process is combining and aggregating several MECOMs into unilevel and multilevel MECOMs.

2.6 Model Validation

In earlier publications [6; 38; 39; 40] we have reported on the progress of the Tactical Real-time Interaction in Distributed EnvironmentTs (TRIDENT) project.

Numerous battle management and emergency response studies have been carried out in which we used every opportunity to test, refine and augment the modeling, measurement, data collection and analysis concepts of TRIDENT [41].

The TRIDENT concepts for analysis and evaluation on aggregated system levels have met high acceptance among the subjects; trained and skilled professionals performing their daily tasks in their accustomed work environment. This is consistent with studies by Rencrantz et al. [42] and Svensson et al. [43] of experience and performance in complex operational settings. However, from time to time it is claimed that reliability and validity of subjective workload ratings are insufficient. To address this critique we designed a study using hormonal response measures, inspired by the results of Svensson et al. [44], who studied workload and performance in military aviation, Zeier, [45] who studied workload and stress reactions in air traffic

controllers, and Holmboe et al. [46], who studied military personnel performing exhausting battle training. The study is described by Norlander [47].

3.0 DISCUSSION

From the study results we could identify a number of particularly interesting causes of mission failure or poor performance. The predominant error modes were:

- Timing of movement and of tactical unit engagement.
- Speed of movement or maneuver, which is especially important in the initial phase of engagement.
- Selection of wrong object. In warfare or crisis response there is high risk for selecting wrong objects, in navigation, in engagements, or in visual contact.

After a retrospective cognitive reliability and error analysis [32] we found that mission failure or poor performance in every case could be attributed to:

- Slow or even collapsed organizational response.
- Ambiguous, missing or insufficiently disseminated, communicated and presented information.
- Equipment malfunction, e.g. power failure or projectile/missile impact.
- Personal factors: inexperience, lack of team training etc.

The results from the hormonal response study suggest three potentially significant mechanisms influencing how the team is able to execute mission control, which consequently also influences mission efficiency [47]:

1. Time-dependant filtering functions like defense and coping mechanisms according to the Cognitive Activation Theory of Stress (CATS).

2. Performance limiting factors due to specific mission and task situation factors and resource requirements.

3. Balance between feedforward and feedback in mission-critical action control.

4.0 CONCLUSIONS

Our theoretical advances and the experimental results validate that cognitive systems analysis of C2 in CMOs is a versatile and effective approach. Cognitive systems analysis facilitates:

1. Identification of limiting factors of a specific individual, unit, system, procedure or mission in CMO.
2. Assessment of the magnitude of influence of these factors on overall tactical performance.
3. Generation and implementation of solutions to improve insufficient capabilities and contribute to successful mission accomplishment.
4. Methodological support to analysis, development and evaluation of complex CMO.
5. Improving training programs for tactical decision making and resource management.

5.0 REFERENCES

- [1] European Defence Agency. (2009) *Future Trends from the Capability Development Plan*.
- [2] Worm, A., Jenvall, J. and Morin, M. (1998). Mission efficiency analysis: evaluating and improving tactical mission performance in high-risk, time-critical operations. *Safety Science*, Amsterdam, The Netherlands, Elsevier Science B.V., Vol. 30, pp.79–98.
- [3] Joint Doctrine Publication 04 Understanding. (JDP-04) The Ministry of Defence Development, Concepts and Doctrine Centre, UK. December 2010.
- [4] Dyer, L., & Shafer, L. (1998). From Human Resource Strategy to Organizational Effectiveness: Lessons from Research on Organizational Agility, *Cornell Working Papers Series*, CAHRS, Working Paper 98-12.
- [5] Rochlin, G. I. (1997). *Trapped in the Net: The Unanticipated Consequences of Computerization*. Princeton, NJ: Princeton University Press.
- [6] Worm, A. (2000a). *On Control and Interaction in Complex Distributed Systems and Environments*. Linköping Studies in Science and Technology, Dissertation No. 664, Linköping University, Linköping, Sweden.
- [7] Flach, J. M. & Kuperman, G. (1998). Victory by Design: War, Information, and Cognitive Systems Engineering. (Report No. AFRL-HE-WP-TR- 1998-0074). Wright-Patterson AFB, OH: Air Force Research Laboratory, Crew System Interface Division.
- [8] Worm, A. (2002). An Integrated Framework for Information-Centered Human-Machine Systems Analysis. *Int. J. Emergency Management*, Vol. 1, No. 2, pp.125–143.
- [9] Rasmussen, J. (1983) Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, pp.257-266.
- [10] Rasmussen, J. (1986) *Information Processing and Human-Machine Interaction: an Approach to Cognitive Engineering*. North-Holland, New York.
- [11] Hollnagel, E. and Woods, D.D. (1983) Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Studies*, Vol. 18, pp.583-600

- [12] Hollnagel, E. (1999) Modelling the controller of a process. *Trans Inst MC*, Vol. 21, Nos. 4/5, pp.163-170.
- [13] Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.
- [14] Brehmer, B. (1992) Dynamic decision-making: human control of complex systems. *Acta Psychologica* 81, pp.211–241.
- [15] Norlander, A. (2009). Representation and Control of Complex Joint Human-Machine Systems: An Information and Sensemaking Perspective. In *Proceedings of The American Society of Naval Engineers Human Systems Integration Symposium 2009*, Alexandria, VA: The American Society of Naval Engineers.
- [16] Moffat, J., 2002, *Complexity Theory And Network Centric Warfare*. Defence Science And Technology Laboratory, UK.
- [17] Grisogono, A., Ryan, A.J., 2003, Designing Complex Adaptive Systems for Defence. In *Proceedings of the Systems Engineering Test and Evaluation Conference*, Canberra.
- [18] Orhaug, T. (1995). Science of Command and Control - A Discussion of Descriptive and Theoretical Problems within the Domain of Command and Control. (In Swedish). The Swedish National Defence College, Stockholm.
- [19] Brehmer, B. (1988). Organization of decision making in complex systems. In L. P. Goodstein, H. B. Andersen, & S. E. Olesen (Eds.). *Tasks, Errors, and Mental Models*. London: Taylor & Francis.
- [20] Annett, J. (1997). Analysing Team Skills. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.). *Decision Making Under Stress: Emerging Themes and Applications*. Ashgate.
- [21] Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International journal of System Science*, 1, pp. 89-97.
- [22] Kailath, T. (1980). *Linear Systems*. Englewood Cliffs, N. J.: Prentice-Hall.
- [23] TRADOC Pamphlet 525-5-500, *The U.S. Army Commander's Appreciation and Campaign Design*.
- [24] Endsley, M.R. (1995) Towards a theory for situation awareness in dynamic systems. *Human Factors*, Vol. 37, pp.32-64.
- [25] Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Ashgate.
- [26] Kleindorfer, P.R., Kunhreuther, H.C. and Schoemaker, P.J. (1993). *Decision Sciences: an Integrative Perspective*. Cambridge University Press, Cambridge.
- [27] Zachary, W. W., & Ryder, J. M. (1997). Decision Support Systems: Integrating Decision Aiding and Decision Training. In M. Helander, T. K. Landauer, & P. Prabhu. (eds.) *Handbook of Human-Computer Interaction*, 2nd ed. Elsevier Science B.V.
- [28] Klein, G. A. (1993a). *Naturalistic Decision Making - Implications for Design*. State-of-the-Art Report. Dayton, OH: Crew Systems Ergonomics Information Analysis Center, Wright-Patterson Air Force Base.
- [29] Klein, G. A. (1993b). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.
- [30] Zsombok, C. E., & Klein, G. A. (Eds.). (1997). *Naturalistic Decision Making*. Mahwah, NJ: Earlbaum.

- [31] Klein, G. A., & Woods, D. D. (1993). Conclusions: Decision making in action. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.
- [32] Worm, A. (2001). Tactical Mission Analysis by means of Naturalistic Decision Making and Cognitive Systems Engineering. In E. Salas & G. A. Klein (Eds.), *Linking Expertise and Naturalistic Decision Making*, pp.407-431. Mahwah, NJ: Earlbaum. ISBN: 0-8058-3538-5.
- [33] Cannon-Bowers, J. A., Salas, E., & Pruitt, J. S. (1996). Establishing the boundaries of a paradigm for decision-making research. *Human Factors*, 38, pp. 193-205.
- [34] Serfaty, D., & Entin, E. (1997). Team Adaptation and Co-ordination Training. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications*. Ashgate.
- [35] Levine, S., & Ursin, H. (1991). What is stress? In M. R. Brown, G- F. Koob, & C. Rivier (Eds.), *Stress – Neurobiology and Neuroendocrinology*. Marcel Dekker, New York, pp. 3-21.
- [36] Eriksen, H. R., Olf, M., Murison, R., & Ursin, H. (1999). The time dimension in stress responses: relevance for survival and health. *Psychiatry Research*, 85, pp. 39-50.
- [37] Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey and E. Salas (Eds.), *Teams: Their training and performance*, pp. 3-30 Norwood, NJ: Ablex Publishing Corporation.
- [38] Worm, A. (1999a). Mission Efficiency Analysis of Tactical Joint Cognitive Systems. In *Proceedings of the NATO International Symposium on Modelling and Analysis of Command and Control*, Paris, France.
- [39] Worm, A. (1999b). Evaluating Tactical Real-time Interaction in Multi-agent, Dynamic, Hazardous, High-stake Operations. In *Proceedings of The Human Factors and Ergonomics Society 43rd Annual Meeting*, The Human Factors Society, Santa Monica, CA., USA.
- [40] Worm, A., Jenvald, J. and Morin, M. (1998). Mission efficiency analysis: evaluating and improving tactical mission performance in high-risk, time-critical operations. *Safety Science*, Elsevier Science B.V., The Netherlands, Vol. 30, pp.79-98.
- [41] Worm, A. (2001) On systems analysis and performance assessment in complex, high-risk work environments. *Int. J. Risk Assessment and Management*, Vol. 2, Nos. 3/4, pp.276–287.
- [42] Rencrantz, C., Lindoff, J., Svensson, E., Norlander, A., & Berggren, P. (2006). Interoperability and methodology study in an operative command and control setting. FOI-R-2040—SE, ISSN 1650-1942 (in Swedish).
- [43] Svensson, E., Rencrantz, C., Lindoff, J., Berggren, P., & Norlander, A. (2006, October 16-20). Dynamic measures for performance assessment in complex environments. In *Proceedings of The Human Factors and Ergonomics Society 54th Annual Meeting*, Santa Monica, CA: The Human Factors Society.

- [44] Svensson, E., Angelborg-Thandertz, M., & Sjöberg, L. (1993). Mission Challenge, Mental Workload and Performance in Military Aviation. *Aviation, Space and Environmental Medicine*, Nov., pp. 985-991.
- [45] Zeier, H. (1994). Workload and psychophysiological stress reactions in air traffic controllers. *Ergonomics*, 37, pp.525-539.
- [46] Holmboe, J., Bell, H., & Norman, N. (1975). Urinary Excretion of Catecholamines and Steroids in Military Cadets Exposed to Prolonged Stress. *Försvarsmedicin*, 11, p. 183.
- [47] Norlander, A. (2010). Analysis of Tactical Missions: Integrating Systems Theory, Cognitive Systems Engineering and Psychophysiology. In Proceedings of The Human Factors and Ergonomics Society 54th Annual Meeting, Santa Monica, CA: The Human Factors Society.

6.0 ACKNOWLEDGMENTS

This work was supported by the Swedish Armed Forces, the Swedish National Defence College, the Swedish Defence Research Agency and the Swedish National Foundation for Strategic Research in collaboration with Linköping Institute of Technology and Stockholm Royal Institute of Technology.

1.12 Colonel Blotto Games and Lancaster's Equations: A Novel Military Modeling Combination

Colonel Blotto Games and Lancaster's Equations: A Novel Military Modeling Combination MODSIM World Paper 2011

Andrew J. Collins

Virginia Modeling, Analysis and Simulation Center (VMASC), Old Dominion University
ajcollin@odu.edu

Patrick T. Hester

Dept. of Engineering Management and Systems Engineering, Old Dominion University
pthester@odu.edu

Abstract. Military strategists face a difficult task when engaged in a battle against an adversarial force. They have to predict both what tactics their opponent will employ and the outcomes of any resultant conflicts in order to make the best decision about their actions. Game theory has been the dominant technique used by analysts to investigate the possible actions that an enemy will employ. Traditional game theory can be augmented by use of Lanchester equations, a set of differential equations used to determine the outcome of a conflict. This paper demonstrates a novel combination of game theory and Lanchester equations using Colonel Blotto games. Colonel Blotto games, which are one of the oldest applications of game theory to the military domain, look at the allocation of troops and resources when fighting across multiple areas of operation. This paper demonstrates that employing Lanchester equations within a game overcomes some of practical problems faced when applying game theory.

1.0 INTRODUCTION

Military strategists face a difficult task when engaged in battle against an unknown adversarial force. They must allocate their resources in the most efficient manner possible so as to maximize their chance of defeating their opponent, while being mindful of their finite supply of soldiers, weapons, ammunition, etc. This task is not straightforward, nor trivial. To assist in this effort, military strategists typically employ computational approaches to help them evaluate potential battle outcomes and positively influence the overall outcome of a particular engagement. Mathematical models provide the theoretical foundation for these computational approaches. These models, while they can be effective, fall into two camps: 1) those that have a basis in Cold War Era tactics which focus on large scale conflicts with massive weaponry, or 2) newer, agent-based models which are computationally expensive and, often, unable to be validated [1]. This paper introduces two of these mathematical models for battle prediction, namely: Lanchester equations and game theoretic models. The theoretical foundations of game theory and then Lanchester models

are discussed. Then, a comprehensive example is provided which showcases the authors' novel approach for combining these techniques used together within the Colonel Blotto problem framework.

2.0 GAME THEORY

Game theory is the study of decision problems involving more than one intelligent agent, or player, and it is used to model how these sophisticated agents interact. The term 'game theory' comes from the application of 'games of strategy' to economic problems by John Von Neumann and Oskar Morgenstern in their seminal book *Theory of Games and Economic Behavior* [2]. Game theory is concerned with determining the *best* way to play a game for any given set of rules. Defining what is meant by best is a non-trivial undertaking because any strategy used by a player must take into account the other player's strategy as well. Game theory is unconcerned about what are good strategies for playing the game unless that good strategy happens to be the best strategy; this distinguishes game theory from everyday gaming.

Game theory was started when, in the nineteenth century, Antoine Cournot proposed an idea that economists should look at situations where there are only a few competitors [3]. Economists had, until that point, only looked at markets without competition, called "Crusoe on his island", or markets when there was infinite competition, called "Multeity of atoms" [4]. The work by Cournot was virtually ignored until John Von Neumann and Oskar Morgenstern wrote their ground-breaking book during the Second World War [2]. This book became the bedrock of modern Game Theory. Seven years later, John Nash developed his Nash Equilibrium concept [5] which allowed Game Theory to become the useful technique within the modern day modeling community.

Classic applications in the defense realm range from looking at international power struggles in Mesquita and Lalman's War and Reason [6] to whether a victor should be magnanimous after the conflict in Bram's Theory of Moves [7]. Game theory has also been applied to the nuclear arms control negotiations that occurred between the United States of America (USA) and Union of Soviet Socialist Republics (USSR) during the Cold War [8].

2.1 Game Representation

Game theory attempts to mathematically capture behavior in strategic situations, or games, in which an individual's success in making choices depends on the choices of others. There are two standard forms with which game theory attempts to display the overall game under consideration; they are: *normal* and *extensive*. Normal-form games use a payoff matrix, for example see Figure 1. Extensive-form games use a game tree, for example see Figure 2.

		RED	
		A	B
BLUE	A	5, 5	0, 9
	B	9, 0	1, 1

Figure 1: Normal-form of a game

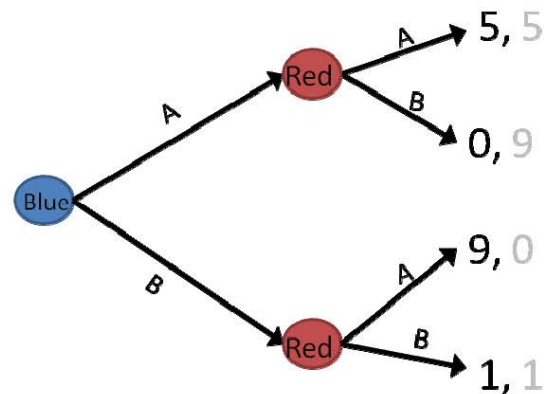


Figure 2: Extensive-form of a game

The rows of a payoff matrix represent the possible actions available to the blue player (BLUE) and the columns represent the possible actions available to the red player (RED). The two values in the cells of the matrix represent that reward or payoff that the players receive. The first number is the payoff to BLUE and the second number is the payoff to RED.

2.2 Nash Equilibrium

The main solution method for game theory is the Nash Equilibrium. It is concerned with the stability of the strategies chosen by the players. A player's strategy is how they choose their action for a game. If a player chooses to go with a single action then this is called a pure strategy. If a player chooses to randomly select an action from several different actions then this is called a mixed strategy. When mixed strategies are used by any of the players, we consider the player's payoff to be the expected payoff.

Given a particular set of strategies for the players it is a Nash Equilibrium if the following statement is true for all players: *Each player does not benefit from changing their current strategy, given the current strategy of the other players*. This does not mean that the players get the maximum payoff available to them within the game but that they gain the highest payoff available under the constraint of the other players' strategies. Some games have multiple Nash Equilibrium solutions.

2.3 Colonel Blotto

One of the oldest applications of game theory to the military domain are Colonel Blotto games. Colonel Blotto games look at the allocation of troops and resources when fighting across multiple areas of operation. The game is based around the simplest assumption that having more troops means that you are more likely to 'win' or control a particular area of operation. The actions that are available for the players are how they allocate their troops to the various areas of operations.

The original paper on Colonel Blotto games was written by Émile Borel, known for his contributions to measure theory, which acts as the foundation for all modern probability theory [9]. Though originally written in 1921, it was translated from French in the 1950's, which was the same time that game theory was becoming popular in the English speaking world due to Von Neumann and Morgenstern's work. It is no consequence that game theory's popularity coincided with the start of the Cold War.

2.4 Criticisms of Game Theory

There are several problems with the application of game theory and a brief summary is given here:

- Payoff determination: It can be difficult to determine what the payoffs for a game should be for each player's actions. It might seem that in some cases the payoffs are obvious, i.e. if the game is about winning money, then the money won or lost should

be the payoff; however, this has been shown in examples like the Ultimatum game not to be the case [15]. As game theory's usefulness is about giving insight, and it is not an exact predictive solution, then the use of approximate payoffs is not a "show-stopper" for its application. In many games the outcome remains the same even with slight changes to the payoff. Given this limitation, the authors see utility in utilizing a Lanchester equation-based approach to determining more sophisticated payoffs.

- Determining the Nash Equilibrium: Games can be very easy to construct but can be difficult to solve. This problem can be seen in the game of chess. Chess is a relatively simple game that has been around for centuries, yet no solution to the game has been found, even though a solution has been shown to exist.

- Mixed strategies: When the solution to a game is a mixed strategy this implies that players should randomly choose between different actions. This might make some sense if the game is repeatedly played, but for one-off games this becomes problematic. Imagine that the mixed strategy says 'play action 'A' 99% of the time and action 'B' 1% of the time'. What would you do in this situation? It can be hard to explain this need to randomize to the player. It is also important that your opponent knows that you are going randomly choose your strategy, but how is this done? Should you phone your opponent and tell them you are rolling a die?

- Rational: Game theory assumes that the players are perfectly rational and infinitely intelligent. This assumption might be fine when highly skilled game players are being modeled with game theory, but is not necessarily the case for everyday people in their everyday lives.

3.0 LANCHESTER EQUATIONS

Lanchester analyzed World War I aircraft engagements and developed a theory to explain why concentration of military effort

in these entanglements was advantageous [10]. This theory was encapsulated as laws, which are low-resolution aggregated models for determining a battle outcome between two opposing forces. For the purposes of this discussion, the two forces considered are denoted as the blue force (the force the analyst is planning for, aiding, etc.) and the red force (the force the analyst is trying to defeat). This model results in a system of differential equations as follows:

$$dB/dt = f_1(R, B...) \quad (1)$$

$$dR/dt = f_2(R, B...) \quad (2)$$

Where B denotes the blue force and R, the red force. This system of differential equations can be solved by substituting unique functions into Eqs. (1) and (2), depending on whether or not the analyst is determining the outcome of an ancient or modern conflict, resulting in Lanchester's Linear and Square Laws, respectively. The linear law was a formulation for ancient warfare, in which one-on-one combat was the only mechanism of engagement. Recognizing the limitation of his linear law to only describe one-on-one combat, Lanchester formulated his square law, which could be utilized to describe many-on-many combat.

3.1 Stochastic Lanchester Systems

The original formulation of Lanchester's models did not include stochastic behavior. That is, all parameters in the model are taken to be deterministic. This is a limiting assumption of the Lanchester models and does not represent reality, as uncertainty abounds, especially in combat situations. To that end, several extensions which amend Lanchester's model to include stochastic behavior have been developed.

Brown [11] developed an approximation to estimating the probability that B blue force members would neutralize R red force members (denoted as $P_{B,R}$).

$$P_{B,R} = \sum_{j=1}^B \frac{(-1)^{B-j} j^{B+R}}{[(B-j)!(R+j)!]} \quad (3)$$

Where j is a variable indice, and all other variables are as before. Equation 3 assumes that each force is depleted by one individual per round. That is, two blue and two red members can be reduced to two blue and one red member or one blue and two red members and not to one blue and one red members (in one battle step). This equation can be used to determine battle outcome probabilities to utilize as outcomes within a game theoretic construct.

3.2 Criticisms of Lanchester Equations

Critics of Lanchester models point out several deficiencies in the original formulation of Lanchester's laws that the authors would be remiss to not mention. Taylor succinctly summarizes them in [12]. Two criticisms were: "Tactical decision processes not considered" and "Battlefield intelligence not considered". Both these criticisms are addressed by coupling Lanchester equations with a game theoretic approach.

Several extensions to the original formulation help to alleviate many of the objections raised. For example, as discussed, the extension proposed by Hester and Tolk [13] allowed for a number of discrete "mini-battles", providing for the discretization of a battle landscape into smaller conflicts.

Despite these shortcomings, the authors believe that the novelty and usefulness of Lanchester models lie in part in their simplicity. The authors do not believe that a simpler, more efficient means of analyzing combat is available to today's modeler. Thus, many have adapted Lanchester's original equations to account for modern warfare scenarios and the authors believe this is a both a valid and useful approach to simple analytical combat modeling, especially when employed in conjunction with the game theoretic environment discussed in the following section.

4.0 MODEL

As mentioned previously, one major drawback in game theory is in the naïve generation of payoffs for games. In an effort to resolve this issue, the authors propose utilizing the battle prediction of Eq. 3 (derived from Lanchester's stochastic square model) to develop the payoffs of the game. A Colonel Blotto game is used as an example and the results are shown in Table 1.

Table 1: Battle Outcome Prediction as Payoffs in a Simple Game

# of BLUE soldiers	# of RED soldiers				
	0	1	2	3	4
0	0, 0	0, 1	0, 1	0, 1	0, 1
1	1, 0	0.5, 0.5	0.167, 0.833	0.042, 0.958	0.008, 0.992
2	1, 0	0.833, 0.167	0.5, 0.5	0.225, 0.775	0.081, 0.919
3	1, 0	0.958, 0.042	0.775, 0.225	0.5, 0.5	0.26, 0.74
4	1, 0	0.992, 0.008	0.919, 0.081	0.74, 0.26	0.5, 0.5

Table 1 represents the normal-form of a game. This game is where the players have to allocate up to four soldiers to attack/defend a single guard post. The payoffs for this game are the players' chance of winning the guard post, as calculated using Eq. 3. Unsurprisingly, the Nash Equilibrium of this game is when both players allocate the maximum number of soldiers to the guard post.

Now let us consider the case when there are two guard posts to be attacked / defended. The blue player will allocate soldiers to the front guard post and the remainder will guard the rear post. Similarly, the red player will allocate soldiers to attack the front guard post and the remainder will attack the rear. Once force distributions are chosen via the Lanchester-assisted game theoretic model, an overall battle success probability can be calculated. Results for the analysis of this associated four person battle, and associated force division are shown in Table 2. In this case, the feasible scenarios for this battle are shown (with

redundant scenarios eliminated), with the probability of the blue team winning in both a series and parallel (payoffs given in grey brackets) scenario also provided. Only one player's payoff needs to be presented in the table as the game is sum-zero (meaning that any red gain represents a blue loss). The actions represent the number of soldiers allocated to the front guard post with the remaining being allocated to the rear guard post.

Table 2: Battle Outcome Prediction as Payoffs in a Colonel Blotto Game.

# of BLUE soldiers at front post	# of RED soldiers at front post				
	0	1	2	3	4
0	0.5 (0)	0.74 (0)	0.919 (0)	0.992 (0)	1 (0)
1	1 (0.26)	0.75 (0.25)	0.813 (0.129)	0.96 (0.04)	1 (0.08)
2	1 (0.081)	0.871 (0.188)	0.75 (0.25)	0.871 (0.188)	1 (0.081)
3	1 (0.008)	0.96 (0.04)	0.83 (0.129)	0.75 (0.25)	1 (0.26)
4	1 (0)	0.992 (0)	0.919 (0)	0.74 (0)	0.5 (0)

The Nash Equilibrium solution to this game is not obvious, unlike the single guard post version. The Gambit software package [14] was used to find the Nash Equilibrium of these games. In the game where only one guard post had to be defended by BLUE, a mixed strategy was found. Their percentage selections of the five actions for the blue player were: (27%, 15%, 16%, 15%, 27%) and for the red player (5%, 36%, 17%, 38%, 5%). This results in an expected payoff for the blue player of 0.86. To gain some understanding to what this payoff means, if the blue player had chosen just to randomly allocate all their soldiers to either of the guard posts, then his or her payoff would have been 0.75.

The mixed strategy for the game when BLUE must win both guard posts results in a payoff of 0.14. Again this game was solved by Gambit and the strategies are (0, 50%, 9%, 41%, 0) for the blue player and (45%, 0, 1%, 54%, 0) for the red player. The reason for these strategy allocations is not

immediately obvious and further investigation would be required.

5.0 CONCLUSION

The authors believe this novel approach is useful when compared with the traditional Colonel Blotto approach, whereby payoffs are not analytically derived. This approach affords the battle planner the ability to understand battle results probabilistically, and it can be extended to more complicated battle scenarios (e.g. more combatants, more than two battles).

Military strategists are able to use Lanchester equations and game theory as tools to enable their decision making when planning for a battle against adversarial forces. Using Lanchester equations provides insights into the possible outcomes from a particular conflict, especially the expected attrition levels. Game theory gives insights to what strategies might be employed by an adversarial force and what is the best response to these strategies. A key strength of both of the analytical techniques is their simplicity, which helps any decision-maker gain clear insights into the problem being investigated by presenting the solution in a concise manner.

There is a temptation by modelers to increase the complexity of their models and simulations due to cheaply available computing power. As we saw in the example combining Lanchester equations and game theory given in the previous section, extra complexity within the model produces extra complexity within its solutions. This extra complexity of the solution is too often ignored by only presenting the simple statistics of the results to the decision-makers. The authors would argue that unless the richness of results from a complex model is going to be explored for further utility, then the modeler should utilize simple models to ensure they remain fully aware of what the results imply and gain the insight from that knowledge. Complexity can bring another problem during the modeling stage, as it is easy to

construct games that cannot currently be solved – like chess.

As with all modeling methodologies, it is important to remember why you are using a particular technique. Both Lanchester models and game theory represent useful methodologies that can be imbedded into a larger simulation to solve the particular problems that they are designed for, i.e., attrition determination and strategy choice. It is in this embedded usage that the authors envision the true utility of this novel approach.

6.0 REFERENCES

- [1] Champagne, L.E., & Hill, R.R. (2007). Agent-model validation based on historical data. *Proceedings of the Winter Simulation Conference*, pp: 1223-1231, IEEE Computer Society, Washington, DC.
- [2] von Neumann, J., and O. Morgenstern (1944). *Theory of games and economic behaviour*. Princeton University Press.
- [3] Cournot, A.A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses* (Researches into the Mathematical Principles of the Theory of Wealth). Paris: Hachette.
- [4] Eatwell, J., M. Milgate, and P. Newman, eds. (1987). *The new palgrave: Game Theory*. Macmillan Press.
- [5] Nash, John. 1951. Non-cooperative games. *The Annals of Mathematics* 54, no. 2: 286–295.
- [6] Mesquita, B.B. , and D. Lalman (1994). *War and Reason: Domestic and International Imperatives*. Yale University Press.
- [7] Brams, S.J. (1993). *Theory of Moves*. Cambridge University Press.
- [8] Aumann, R.J. (1995). *Repeated Games with Incomplete Information*. The MIT Press, 1995.

- [9] Borel, E. 1953. The Theory of Play and Integral Equations with Skew Symmetric Kernels. *Econometrica* 21, no. 1 (January): 97-100.
- [10] Lanchester, F. W. (1914). "Aircraft in Warfare: The Dawn of the Fourth Arm — No. V, the Principle of Concentration." *Engineering*, 98: 422-423.
- [11] Brown, R.H. (1963). Theory of combat: the probability of winning. *Operations Research*, 11(3), 418-425.
- [12] Taylor, J.G. (1983). *Lanchester Models of Warfare, Volume I*. Arlington, VA: Operations Research Society of America.
- [13] Hester, P.T. & Tolk, A. (2009). Using Lanchester equations for sequential battle prediction enabling better military decision support. *Int. J. Intelligent Defence Support Systems*, 2(2), 76-90.
- [14] McKelvey, R.D., A.M. McLennan, and T.L. Turocy (2007). *Gambit: Software Tools for Game Theory - Version 0.2007.01.30*.
<http://econweb.tamu.edu/gambit>.
- [15] Werner, G.; Schmittberger, R. and Schwarze, B. (1982); An Experimental Analysis of Ultimatum Bargaining; *Journal of Economic Behavior and Organization*, 3, no. 4: 367-388.

1.13 Mission Assurance Modeling and Simulation: A Cyber Security Roadmap

Mission Assurance Modeling and Simulation: A Cyber Security Roadmap

Gerald "Jay" Gendron
gerald.gendron@gmail.com

David Roberts
Donold Poole, CISSP
Anna Aquino
SimIS Inc.
Portsmouth, VA 23704
{david.roberts, donold.poole, anna.aquino} @simisinc.com

Abstract. This paper proposes a cyber security modeling and simulation roadmap to enhance mission assurance governance and establish risk reduction processes within constrained budgets. The term mission assurance stems from risk management work by Carnegie Mellon's Software Engineering Institute in the late 1990s. By 2010, the Defense Information Systems Agency revised its cyber strategy and established the Program Executive Officer-Mission Assurance. This highlights a shift from simply protecting data to balancing risk and begins a necessary dialogue to establish a cyber security roadmap. The Military Operations Research Society has recommended a cyber community of practice, recognizing there are too few professionals having both cyber and analytic experience. The authors characterize the limited body of knowledge in this symbiotic relationship. This paper identifies operational and research requirements for mission assurance M&S supporting defense and homeland security. M&S techniques are needed for enterprise oversight of cyber investments, test and evaluation, policy, training, and analysis.

1.0 INTRODUCTION

*It was the best of times, it was the
worst of times, it was the age of
wisdom, it was the age of
foolishness, it was the epoch of
belief, it was the epoch of
incredulity...*

~Charles Dickens

This opening passage from Charles Dickens' 1859 classic *A Tale of Two Cities* seems a fitting statement of cyber space in the early twenty-first century. How timeless Dickens' sense of blessings and curses – and applicable to the modern-day Information Age. This sense of paradox motivated the authors to consider the topic of cyber security and launch a discussion about the balance that must be established in the realm of cyber operations. The objective is to blaze a trail towards a Cyber Security Roadmap, specifically a roadmap harnessing the talents of both operators and

operations analysts in a collaborative partnership and develop concrete actions which address these questions: how can models and simulations add value to cyber security as they have with test and evaluation, design, and decision making?

The term *mission assurance* was found in the literature as early as the early 1990s out of Carnegie Mellon's Software Engineering Institute. Their seminal work in this area used the term within the context of risk management. In the estimation of the authors, this approach leaning on risk management is a timeless rendering of the term mission assurance and is noted here for its sound approach. Unfortunately, within the context of information assurance, risk management is often misinterpreted as risk *avoidance*. Highlighting the importance of word choice, the Defense Information Systems Agency revised its cyber strategy. As part of reorganization, the Agency established the office of the Program Executive Officer – Mission Assurance.

The authors envision a Cyber Security Roadmap which addresses two key characteristics. The first is governance.

How can cyber security policies and practices be established to best articulate operational requirements which could benefit from analytical collaboration?

Second is the treatment of risk management.

What actions can facilitate a shift in information assurance culture to best integrate the treatment of risk management into the enterprise?

For instance, the National Institute of Standards and Technology (NIST) Special Publication 800-39 entitled "Integrated Enterprise-Wide Risk Management" is a recent, cutting-edge work dedicated to shifting the culture with regard to better understanding the proper tenets of risk management [1].

2.0 LITERATURE REVIEW

The concept of mission assurance is emerging as a shift in culture and perspective from information assurance. Mission assurance provided the authors with a compelling vision of the need for further research in this area. Their goal was to review the body of unclassified literature for work in the area of modeling and simulation supporting information security. This section summarizes the results of a modest literature review using sources from both academia and industry.

The methodology used to research the academic body of knowledge was by means of the EBSCOhost search engine made available through National Defense University in Washington, D.C. Less than a dozen references were identified which met the Boolean search parameters for both

"modeling and simulation" as well as "information security". Many of the resultant articles were related to the simulation of cryptologic key algorithms and not directly applicable to this paper. These scant results led the authors to carefully consider their search techniques to ensure a body of work was not being overlooked. Results could not be improved by using variations on the search terms. It was later learned in researching open source commercial work that the field of modeling and simulation in support of information security is either fairly nascent or classified.

Three peer-reviewed journal articles were most interesting in their treatment of analysis within cyber security. Sahinoglu [2] presents his work on the need for quantitative methods to improve risk management techniques in software intensive systems. He presents a Security Meter Model based on a relatively classic representation of risk:

$$RR = \text{Vulnerability} \times \text{Threat} \times \text{LCM} \quad (1)$$

where RR is the Residual Risk and LCM represents a Lack of Countermeasure. Using Bayesian probability methods, Sahinoglu [2] determines the probability data for various threats pertaining to sets of vulnerabilities. His objective is to quantify enhanced risk postures relative to the incremental costs of countermeasures, "guiding the security manager as to how one can move to a more advantageous state from the present" ([2], p. 1257). In essence, Sahinoglu's work establishes an objective and quantitative cost-benefit approach to capital planning of information security investments.

An article by Kumar, Park, and Subramaniam [3] entitled "Understanding the Value of Countermeasure Portfolios in Information Security Systems" presents an analysis which complements the work by Sahinoglu. Their work focuses on the

challenges organizations face to properly “understand the economic consequences of security attacks relative to the ISSC [information system security countermeasures] portfolio implemented” ([3], p. 241). Kumar, Park, and Subramaniam make three key contributions to the body of literature:

- developing a simulation to integrate threat, prevention, and recovery;
- investigating how to better model interactions among business, threat, and countermeasure parameters;
- recognizing an information security portfolio is dynamic and discusses models and simulations based on financial asset valuation.

Kumar, Park, and Subramaniam [3] – like Sahinoglu [2] – offer knowledge for potential use in a Cyber Security Roadmap. Finally, an article by Ying, Lin, Lee, and Lin [4] investigates some sophisticated modeling based on neural networks to classify an increasing threat in today’s cyber environment – unwanted emails. It focuses on filtering spam e-mail. Their work can be used more generally to model the identification of other types of unwanted e-mail such as the increasing difficulties in interdicting e-mails used to perpetrate spear-phishing attacks.

In addition to more traditional sources from academia, the Google search engine provided the authors with a glimpse into industry’s body of knowledge with regard to modeling and simulation in support of information security. Due to the author’s desire to maintain research at the unclassified level, the results of the literature review were relatively meager. This could be due to limited work in this field, lack of transparency in the interest of closely held corporate research, or the potentially classified nature of this work. However, there were two professional

forums that have had recent professional events including industry members that pierced the subject of interest. Most notably was the conference held in March 2011 under the auspices of the Military Operations Research Society entitled “Mission Assurance: Analysis for Cyber Operations” [5]. This special meeting was a follow up event to a 2008 Cyber Analysis Workshop held in Reston, Virginia [6]. Additionally, a second forum of note was the 16th International Command and Control Research and Technology Symposium held in June 2011 by the Command and Control Research Program out of the Office of the Assistant Secretary of Defense (NII). In particular, a Cyber track was added to this year’s Symposium and highlighted the need for future research in this area.

3.0 “GAME CHANGERS” FOR CYBER ANALYSIS

This section provides a summary of the two events to include event objectives, key findings, and recommendations. These events publically identified a need to develop solutions for problems associated with analysis of cyber security.

3.1 Military Operations Research Society Special Meeting

The Military Operations Research Society (MORS) is a professional organization dedicated to understanding the needs of military users and encourages operations research to support military operations [7]. When particularly broad-reaching and complex operational problems emerge within the defense sector, MORS has sometimes called for a Special Meeting on the subject to bring together experts from academia, industry, and the government to begin a dialogue on the issue. In March 2011, conditions warranted the establishment of a Special Meeting entitled “Mission Assurance: Analysis for Cyber

Operations". From March 21 – 24, 2011, a diverse body of people including operational analysts and information technology specialists convened at Southwest Research Institute in San Antonio, Texas. Their express purpose was to establish cross-functional working groups to consider four primary components of cyber security: (a) situational awareness; (b) establishing and extending networks; (c) operating and defending networks; and (d) the application of cyber force. The Special Meeting had four objectives:

- Enhance understanding of cyber threat among participants;
- Advance analytical methods to support cyber operations;
- Facilitate conversation among operators, analysts, and cyber customers to improve mission assurance;
- Write an unclassified report with classified appendices articulating analytic techniques and posit recommendations to improve cyber operations and mission assurance [6].

The four working groups generated many findings and recommendations which are too numerous to completely discuss in this paper. The authors encourage interested readers to visit the MORS website [6] to access and read the Special Meeting agenda, working group presentations and the final briefing synthesized from those working groups. The authors would like to highlight a few key findings and their associated recommendations as they relate to Mission Assurance Modeling and Simulation and a Cyber Security Roadmap. First among these are findings related to operationally define and communicate the problems [5]. Terms like Mission Assurance and Cyber Domain must be defined among various functional groups from operators to analysts. Likewise, consistent lexicon and

metrics currently do not exist to forge discussions among operational and analytical professionals. MORS recommendations include:

- establishing tactics, techniques, and procedures as well as doctrine;
- refining and using a "Joint Staff Cyber Lexicon" ([5], slide 17);
- articulating cyber networks, attacks, and defenses in operational planning terms and the Joint Operation Planning Process.

Secondly, many operational support-oriented findings were presented ranging from training to organizing and equipping operational forces. Most notable among the findings was the lack of mutual understanding between cyber operators and the analytic community – the former discuss the problem in qualitative terms whereas the latter desire direct metrics. Cyber operators typically lack knowledge of how operational analysis can assist operations and the analytic community currently has few practitioners interested in cyber warfare. Additionally, cyber warfare is in the nascent stage of development without an organized body of knowledge and subject to inadequate or inefficient practices, organizational constructs, and policies. MORS recommendations include:

- establishing a MORS Cyber Analysis Community of Practice;
- nurturing more work in the area of cyber analysis;
- communicating how analysis could aid the cyber operations force.

This is as much a matter of emphasis and leadership as it may be related to funding and training.

Finally, there are future research areas articulated in the MORS findings and recommendations. Numerous operations

research techniques such as Statistical Process Control, Design of Experiments, Neural Networks, optimization methods, and Decision Analysis tools can be employed to model and analyze the networks themselves as well as cyber enablers like manpower and portfolio prioritization. The forum also found an unsatisfactory understanding of cyber and virtual threats. MORS recommendations include:

- developing force-on-force models that account for cyber effects and activities;
- building a bibliography of existing literature on this area;
- establishing more rigorous analysis of the threat;
- highlighting the need for leadership in this area to make this a priority across the Services and instilling a *pull* from the cyber force.

3.2 Command and Control Research Program Symposium

The Command and Control Research Program (CCRP) is a group within the Office of the Assistant Secretary of Defense (NII) dedicated to improve the state of command and control while increasing the understanding of implications of the Information Age on national security [8]. Additionally, it sponsors an annual symposium called the International Command and Control Research and Technology Symposium (ICCRTS). From June 21-23, 2011, the 16th ICCRTS convened in Québec City, Québec. Participants from over 20 countries included members of academia, industry, and government. This symposium marked the addition of a new track on Cyberspace Management. Proceedings from this event are available online [9]. Presentations for this track were diverse in focus and very professional. There was also ample discussion on how this track may and

should evolve. The track chair noted a desire for papers in next year's event addressing foundational elements of this subject area such as terms of reference, understanding differences between cyber domains and functions, and identifying research needs in the cyber arena.

In addition to this new track, the CCRP Director, Dr. David Alberts, presented summary findings of a draft paper stemming from CCRP workgroup SAS-065 on the NATO NEC C2 Maturity Model. The goal of workgroup SAS-065 was "to create a NATO NEC C2 Maturity Model (N2C2M2) to facilitate the exploration of network-enabled command and control approaches and capabilities in a coalition context" ([10], p. 1). The N2C2M2 includes metrics to help organizations assess their capabilities relative to command and control. SAS-065 was formed in 2006 before the emergence of US Cyber Command. Nonetheless, additional research is indicative of the agility of this model for contemporary issues like cyber security. Noteworthy of the Alberts presentation as it relates to the subject of this paper are his findings that cyber can and should be analyzed in further detail.

4.0 CYBER SECURITY ROADMAP

Based on information in the literature review, the authors propose that the operational and analytic community may best be served through their collaboration in developing and concurring on a Cyber Security Roadmap. This paper proposes a broadly defined Roadmap as a first enumeration of a more detailed and necessary one to be developed in the near term. Its primary objective is to stimulate discussion and collaboration among the various communities of interest which impact or are impacted by cyber security issues. This first cut certainly does not get everything right, but it serves as progress towards a practical end. This initial roadmap touches on two key elements:

operational requirements and research requirements. It currently does not address a third important element which is the temporal aspects of a more refined Roadmap. This is important to help drive decision and funding cycles of the operational and research requirements. Temporal aspects are best addressed by a collaborative team of stakeholders.

4.1 Operational Requirements

United States Cyber Command has the mission to develop and staff issues related to the war fighting functions of offensive and defensive cyber operations. In addition to cyber operations, there are four broad areas of operational support where a Cyber Security Roadmap could offer tangible, strategic end states to guide the operational and analytic communities:

- Cyber Policy
- Investment Strategy
- Test and Evaluation
- Training Development

Briefly touching on each area, *policy makers* may benefit from modeling and simulation to forecast and evaluate policy frameworks against various operational environments and national strategies. This could provide insights as to the impact of various policy measures on cyber operations. Likewise, alternative *investment strategies* could be investigated through modeling and simulation to identify gaps and seams in cyber security portfolios with respect to threats, vulnerabilities, and countermeasures.

The *test and evaluation* community has long relied on the capabilities inherent in modeling and simulation to accomplish assessments more effectively under constrained resources. A Cyber Security Roadmap should articulate a vision of effectively conducting test and evaluation on weapon systems as cyber operations

become more mature. For instance, what does cyberspace *look like*? Perhaps we should embark on an era of great explorers like Vespucci and Columbus. These Renaissance men were explorers, navigators, and cartographers – making the unknown world visible through mapping. Modern-day explorers could define the landscape of cyber space to the benefit of the test and evaluation community. Finally, *training* operators about cyber and with cyber will require a cyber approach. Modeling and simulation should be considered more widely to augment live cyber training events. Not only can simulations generate synthetic training environments, they can be used to collect data to better quantify training effectiveness.

4.2 Research Requirements

In addition to enablers to cyber operations, there are subjects requiring further research with respect to cyber security. The authors have identified three broad areas to categorize research elements of a Cyber Security Roadmap:

- Analytic Capabilities
- Test Infrastructure
- Risk Reduction

4.2.1 Analytic Capabilities

This paper advocates the need for baseline models and simulations for use among partners investigating and analyzing cyber security topics. Potential actions include:

- Capture the current knowledge relevant to the modeling and simulation of information security vulnerabilities and protections, to include incident management. The outcome should identify the needs, translate them into requirements, and provide summary information on resources available to meet the needs and requirements. Since the

landscape of cyber threat environments will be continuously changing, these baseline requirements are intended to be utilized in a suitable systems engineering approach to enable agile and assured systems.

- Decompose the requirements into actionable segments.
- Expand trust-relevant requirements (whether automatically generated or manually specified) beyond control measures described in standards or policy documents.
- Develop trust models and metrics to be used in engineering of system into measurable system attributes and then in testing and assessment.
- Advocate for composable components that could be rapidly reconfigured for addressing unforeseen circumstances.
- Determine ways to treat human behavior representation.
- Prioritize data governance – not all information is original or critical. Channel resources for data protection.
- Determine how to automate threat analysis from several sources to discern cyber attacks and develop remediation strategies.

4.2.2 Test Infrastructure

The term *test and evaluation* is a familiar term representing the ability to assess systems for war fighter effectiveness and suitability. A less common thought with respect to test and evaluation is that its execution relies on a harmoniously balanced interplay among diverse partners. The test and evaluation infrastructure arrayed across the globe helps join these partners. It is also one of the great behind-the-scenes aspects of test and evaluation which is commonly misunderstood. The Test Resource Management Center is

responsible for development of next generation test infrastructure to support test and evaluation in the coming decades. The Cyber Security Roadmap should articulate what cyber security test requirements are looming in the distance so research and development of the requisite test infrastructure will exist to support acquisition program needs. Potential actions include:

- Research systems engineering approaches. This may include an open systems engineering environment that promotes logical and virtual test designs before physical test designs.
- Shift approaches more closely to the optimal model-evaluate-build concept.
- Conduct pilot evaluations using sufficiently mature technologies to perform exploratory development, advanced development, and support actionable segment(s).

4.2.3 Risk Reduction

Risk management concepts are fairly well captured in many bodies of literature. Specifically, the Federal Information Processing Standards publications from the National Institute of Standards and Technology do a very good job of outlining risk management in cyber operations. There exists a need for social research on how well the culture is changing from information assurance to mission assurance. Potential actions include:

- Determining individual and organizational ability to rapidly adapt to unforeseeable threats or opportunities.
- Articulate solution-assurance goals so the resulting systems are trusted, assured, reliable, and interoperable.
- Ability to move from compliance, typically a point-in-time objective, to

continuous monitoring and automated remediation strategy.

5.0 CONCLUSION

It is our choices that show what we truly are, far more than our abilities.

~J.K. Rowling (Harry Potter)

This paper proposes a cyber security roadmap to enhance mission assurance governance and establish risk reduction processes within constrained budgets. The authors begin with a summary of a literature review in this nascent field of study. They also point out many emerging issues in this area which have gained recent attention by key professional groups. A key first step is to identify collaborative opportunities among cyber operators and operations analysts. Professional societies note these two groups have not been the focus of leadership attention and they could work more closely together. The Military Operations Research Society has gone so far as to recommend establishment of a cyber community of practice.

Finally, this paper presents thoughts and recommendations on two key elements of a Cyber Security Roadmap: operational and research requirements. Operationally, this Roadmap suggests four broad areas to focus discussions and end states: cyber policy, investment strategy, test and evaluation, and training development. Each of these four areas stands to benefit from modeling and simulation to aid decision making. Likewise, this Roadmap suggests three broad areas to focus future research in support of cyber security: analytic capabilities, test infrastructure, and risk reduction. These research areas encompass a holistic approach ranging from analytic techniques to the quest for new test

technologies to research on human and cultural impact on the sense of risk.

One key element not addressed in this Cyber Security Roadmap is the temporal aspects. This is important to help drive decision and funding cycles of the operational and research requirements. Temporal aspects are best addressed by a collaborative team of stakeholders.

6.0 REFERENCES

- [1] National Institute of Standards and Technology (NIST). (2010, December). *Integrated enterprise-wide risk management: Organization, mission, and information, system view* (Special Publication 800-39). Gaithersburg, MD: US Department of Commerce.
- [2] Sahinoglu, M. (2008, June). An input-output measurable design for the security meter model to quantify and manage software security risk. *IEEE Transactions on Instrumentation and Measurement*, 57 (6), 1251-1260.
- [3] Kumar, R. L., Park, S., & Subramaniam, C. (2008, Fall). Understanding the value of countermeasure portfolios in information security systems. *Journal of Management Information Systems*, 25 (2), 241-279.
- [4] Ying, K. C., Lin, S. W., Lee, Z. J., & Lin, Y. T. (2010). An ensemble approach applied to classify spam e-mails. *Expert Systems with Applications*, 37, 2197-2201.
- [5] Military Operations Research Society. (2011a, March 24). Synthesis group final report from *Mission Assurance: Analysis for Cyber Operations Conference* [PDF document]. Retrieved from <http://www.mors.org/UserFiles/file/2011>

- %20Cyber%20Assurance/Outbriefs/MORS%20Cyber%202011%20Synthesis%20Final%20Briefing%20(V3).pdf.
- [6] Military Operations Research Society. (2011b). *Terms of reference*. Retrieved from <http://www.mors.org/UserFiles/file/2011%20Cyber%20Assurance/MORS%20Cyber%202011%20Working%20Group%20TOR%20and%20Agenda%208%20Mar.pdf>.
 - [7] Military Operations Research Society. (n.d.). *MORS goals and code of ethics*. Retrieved from <http://www.mors.org/about/code.aspx>.
 - [8] Command and Control Research Program. (n.d.). *About the program*. Retrieved from http://www.dodccrp.org/html4/about_main.html.
 - [9] International Command and Control Research and Technology Symposium. (n.d.). *Proceedings from 16th international command and control research and technology symposium*. Washington, D.C. Retrieved from http://www.dodccrp.org/events/16th_iccrts_2011/post_conference/html/tracks.html.
 - [10] NATO SAS-065 Research Task Group. (n.d.). *NATO NEC C2 maturity model*. Retrieved from http://www.dodccrp.org/sas_files/n2c2m2_final_draft.pdf.

1.14 Framework for Assessing Cost and Technology: An Enterprise Strategy for Modeling and Simulation Based Analysis

Framework for Assessing Cost and Technology: An Enterprise Strategy for Modeling and Simulation Based Analysis

Michael O'Neal
Marine Corps Systems Command

Tommer R. Ender, Daniel C. Browne, Nicholas Bollweg
Georgia Tech Research Institute

C. Justin Pearl and Joe L. Bricio
Navy Surface Warfare Center Dahlgren – Combat Direction Systems Activity

Abstract. Acquisition program managers are often forced to make tough decisions regarding complex systems of systems with limited, or even nonexistent, information. This has led to a number of acquisition programs being cancelled after billions of dollars were spent. Congress has directed the Department of Defense (DoD) to implement cost saving measures, specifically targeting ground combat and armored tactical vehicles. Faced with this challenge, the Marine Corps Systems Command (MCSC) is using modeling and simulation (M&S) tools to support alternatives analysis in fielding future ground and related amphibious vehicles. Furthermore, MCSC has initiated the Framework for Assessing Cost and Technology (FACT) to enable decision makers and analysts to conduct system performance, reliability, and cost trades using an integrated, web browser-based M&S toolset. Data integration is enabled through software architecture development based on a Community of Interest (COI) accepted metadata schema. The metadata standardization facilitates reuse of data and gives the program manager a tool to benchmark their program against similar DoD programs and commercial analogues, leveraging open standards and software where possible. FACT will additionally enable rapid analysis of alternative technology/materiel through the use of surrogate models, or equation regression representations of more complex M&S tools, as proven through several successful implementations discussed in this paper. This process will ultimately reduce program development and life cycle costs, both of which are tenets of effective "should-cost" management.

NOMENCLATURE

AoA	Analysis of Alternatives
CIO	Chief Information Officer
COI	Communities of Interest
DoD	Department of Defense
DoE	Design of Experiments
DIACAP	DoD Information Assurance Certification and Accreditation Process
FACT	Framework for Assessing Cost and Technology
GPGPU	General Purpose Graphics Processing Units
HTTPD	Hypertext Transfer Protocol Daemon
HTTPS	Hypertext Transfer Protocol Secure
M&S	Modeling and Simulation
MCSC	Marine Corps Systems Command
RIA	Rich Internet Application
RDBMS	Relational Database Management System
SPARQL	SPARQL Protocol and Resource Description Framework Query Language
SQL	Structured Query Language
USMC	United States Marine Corps
VV&A	Verification, Validation, & Accreditation

1.0 INTRODUCTION

Program managers must address a number of important questions that drive the success of most any system. How well does the system perform? How reliable is it? How much does it cost? Answering these, and other related questions, is a great challenge for any program manager. Those program managers, however, usually rely on disparate and incomplete information to influence important acquisition related decisions. Important decisions based on insufficient information have led to a number of major program cancellations after much funding has been expended.

Congress has recently directed the Secretary of Defense to assess activities for technology modernization of the ground combat vehicle and armored tactical wheeled vehicle fleets [1]. The need for similar capability was emphasized in a memorandum from the Office of the Secretary of Defense for Acquisition Technology and Logistics, which implemented "will-cost" and "should-cost" management for all major acquisition category programs [2]. The primary goal of

this memorandum was to eliminate cost overruns and begin to deliver programs below budget baselines that are set using independent will-cost estimates. Managers for such programs will be tasked to provide should-cost estimates at major milestone decisions. Clearly, decision makers need to make better informed decisions regarding the economic viability and technical feasibility of major acquisition programs.

Faced with this challenge, the Marine Corps Systems Command (MCSC) is using modeling and simulation (M&S) tools to support analysis of alternatives in fielding the future ground combat vehicle fleet and armored tactical wheeled vehicles. MCSC acquisition programs are currently using M&S to support design processes and reduce program/project risk as part of their decision-making and systems engineering process. M&S is useful during system development, for both new acquisition programs and legacy programs, because it has the potential to:

- Reduce cost, schedule, performance, and supportability risk.
- Reduce the time needed between requirements definition and the delivery of capabilities.
- Reduce infrastructure, resource, and personnel usage.
- Improve the military operational effectiveness and suitability usage.
- Increase the quality, interoperability, and supportability of fielded capabilities.
- Reduce, within M&S itself, duplication of effort and tools.

What is needed is a process that uses M&S to enable rapid trade space and alternatives analysis. Traditionally, program managers use stove piped, serialized, independent processes to reach their acquisition decisions. This proves difficult to determine how changes in technology impact reliability and cost. Therefore, MCSC is initiating the Framework for Assessing Cost and Technology (FACT) to enable decision makers and analysts to conduct system performance, reliability, and cost trades using an integrated, web browser-

based M&S tool set. The primary enabler of this integration will be a software architecture development based on an approved metadata schema [3]. This paper describes a proposed approach to address analysis of alternatives for future USMC combat vehicles. This includes cost, schedule, and performance risk assessment as well as processes that use physics-based non-linear computational models and algorithms, and cost-estimating relationship models to support total lifecycle management of USMC vehicles.

FACT will be developed with the recognition of dependency relationships between technology, resources, and time, leveraging M&S to determine how changes in any of these independent variables will impact the other variables. This ultimately measures the ability of a system to meet its requirements. FACT establishes the required workflow/dataflow dependencies from a standards-based architectural framework to gain a better understanding of the design trade space.

At the core of the FACT process is the use of metadata standardization. The metadata standardization facilitates reuse of data and gives the program manager a tool to benchmark their program against similar DoD programs and commercial analogues. FACT will additionally enable rapid analysis of alternative technology/materiel, reducing development of program lifecycle costs, both of which are tenets of effective should-cost management. The proposed process will integrate the data retained in multiple disparate databases, "cylinders of excellence," into a decision support toolset to permit a program office to conduct technology trade space analysis. Analysis of performance versus cost and/or reliability can be determined in a manner that shows the dependency and interaction of changes.

2.0 DISCUSSION

The enterprise level M&S-based strategy discussed in this paper is driven by several key DoD guidance documents, such as the DoD Net-Centric Data Strategy [3],

which describes a vision for a net-centric environment and includes a number of “data goals” for achieving that vision. Additional direction on implementation of the DoD Data Strategy by the DoD CIO is contained in a directive describing “Data Sharing in a Net-Centric Department of Defense” [4]. These documents, along with the “Strategic Vision for DoD Modeling and Simulation” [5] form the basis of the FACT M&S enterprise data strategy, and contain approaches to achieving the overarching Data Goals for the Department given in Table 1 in relation to the M&S-specific Goals that support achieving the DoD “Strategic Vision for M&S”. The strategy proposed in this paper ascribes to the M&S Strategic Vision Goals and supports the broader context to foster community-wide discoverability, coordination, interoperability, and reuse of M&S generated data.

2.1 Architecture and Data Guidance Overview

The enterprise data strategy discussed in this paper contains five primary Data Goals as discussed below.

Data Goal #1: Leverage DoD Net-Centric and M&S Architectural and Data standards to establish a FACT Architecture to allow for discoverable and sharable data and services. DoD discovery metadata will be used to make M&S generated data visible to the DoD level enterprise. M&S COI-specific discovery metadata will be enhanced for functional users. M&S data will be shared through DoD and M&S enterprise catalogs and other discovery tools. M&S data will be accessible to tools and applications through the adoption of structural metadata, data standards, and standardized data models. M&S data will be made understandable and interoperable through the use of DoD, international, and M&S COI-specific ontologies and semantic metadata.

Data Goal #2: FACT Architecture will allow for a services oriented approach to enable “easy” access and new functionality. The FACT architecture

will be conceived in such a way that the future addition of new M&S functionality will be straightforward. Data schemas and service interaction protocols will be designed and implemented to create an open systems architecture, one which promotes the integration of new services and capabilities as needs arise. Taking a modular approach in establishing the software and data architecture will help simplify the process of adding to the architecture in future iterations.

Data Goal #3: The FACT architecture will allow for effective configuration management and promote visibility of services and data via metadata standards. The base software architecture will include version control software for effective and structured configuration management. Data will be stored and tagged using DoD metadata standards to maintain commonality with existing DoD data catalogs, enhance visibility to outside entities and services, and support the DoD net-centric data goal.

Data Goal #4: FACT data sources and services will provide visibility and pedigree according to verification, validation, and accreditation (VV&A) best practices. Sources of FACT data will be made available and visible to users and inquirers to promote data sharing among DoD entities. Quality and origin of data will be verified, validated, and made visible in order to establish credibility of FACT’s cost and performance analysis results.

Data Goal #5: The FACT architecture will contain the necessary data/information that will enable users to get educated on capabilities/limitations and technical foundations of data and services. The FACT architecture will be designed to be accessible and approachable by users, without regard to their level of technical knowledge or background. This drives the need for documentation that clearly describes functions and capabilities, as well as the limitations of those capabilities. In this way, expectations of users and potential collaborators will be managed.

Table 1. Mapping of DoD Net-Centric Data Goals, M&S Strategic Vision Goals and FACT Architecture and Data Guidance.

DoD Net-Centric Data Goals [3][4]	M&S Strategic Vision Goals [5]	FACT Architecture and Data Guidance
• Make data visible. • Make data accessible. • Enable data to be understandable. • Enable data to be trusted. • Support data interoperability. • Be responsive to user needs. • Institutionalize data management.	• Shared accessible standards, architectures, networks and environments. • Policies at the enterprise level that promote the use of common capabilities and minimize unnecessary duplication. • Management processes that promote sharing, reuse, and cost effective development of M&S tools and data. • Tools and authoritative data that are available to credibly support the full range of DoD interests. • Well-trained workforce.	• Leverage DoD Net-Centric Architectural and M&S data standards to establish an architecture that allows for discoverable and sharable FACT data and services. • FACT Architecture to allow for a services oriented approach to allow for easy access and evolutionary approach to new functionality. • FACT Architecture will allow for effective configuration management and promote visibility of FACT services and data via metadata standards. • FACT Data Sources and Services to provide visibility and pedigree according to VV&A best practices. • FACT Architecture to contain the necessary data/information that will enable users to get educated on capabilities/limitations and technical foundations related to FACT data and services.

A core characteristic of the FACT architecture is that it must be open, services-oriented, and protocol-driven. Data interfaces must be well-defined; M&S modules and functionality must be designed around the interfaces they implement, and not the other way around. The supporting software frameworks that the FACT web application and data services will be built upon must be versatile, robust, and use common interface standards. The use of open-source tools is being widely adopted across the DoD [7], and its utilization for development of the FACT's computational engine and user interface generation software is in alignment with the needs discussed in this section. This will further help achieve the goal of a tool that is

modular, scalable, and agile in response to possible future modification and expansion of capability.

2.2 Architecture Considerations

Scalability. An architecture must be developed that will easily scale to any desired level of future expansion. Scalability is affected by every piece of the architecture: hardware, software, and the underlying software frameworks. Hardware must be purchased and configured in a modular manner so that it can be easily upgraded, without changing other system components or characteristics. Software must be able to utilize new hardware resources and translate them efficiently into improved system performance, and be

modular in design, so that new functionality can be added in the future with minimal modification to existing code.

Performance. The approach needs to provide near real-time performance throughout all possible stages of use. To ensure this, a number of different variables must be considered such as network traffic, location of data sources, hardware specifications and capability, and efficient use of resources by system software. The user experience must not be diminished below an acceptable level, even under less than ideal conditions (i.e. heavy network traffic).

Persistence. Results from past model analyses must be able to be retained, and then retrieved from the system at a later date. This must be taken into account when choosing a data store, as well as the hardware used to store the data. Besides common relational databases, semantic stores should also be considered as they may prove superior in representing the necessary data constructs. With ever-increasing amounts of data being stored, the system should be designed so that storage expansion is trivial.

Security. The entire design and engineering process must be performed with consideration to DoD security and information assurance standards. Incomplete knowledge of these requirements in the initial stages of planning could result in considerable wasted effort, and greatly increase cost and schedule risk. Additional issues regarding security classification related accreditation have a great impact on this effort; the DIACAP, described in detail in DoD Instruction 851.01 will be adhered to ensure rapid certification and accreditation [8].

Data homogenization. One of the core objectives of the FACT is to be a repository for the storage of different data models, representing real world systems, and a center for the cost and performance analysis within the trade space of those models. Accomplishing this requires the assimilation of various types of data from disparate sources, in different formats,

requiring data translation. It is further likely that FACT will pull from data that may be queried, but not downloaded; in this case, FACT data models may be partially or entirely federated. This carries with it performance implications, as well as the necessity to accommodate the static nature of the database being federated, such as with a unique set of queries.

Computational engine. While the earliest prototypes of the FACT will utilize a homogenous modeling approach, selection of an engine that provides heterogeneous model interfaces allows for future extensibility. Among the core capabilities of such a system should be data conditioning, such as to prevent errors introduced through unit of measure mismatch and range incompatibilities. For example, inclusion of existing engineering M&S codes which are available for execution on an application server should be supported.

Collaboration. The FACT is intended to be a collaborative tool, equally accessible to government, academia, industry, and related partners. Care must be taken to ensure that data models can be accessed and analyzed concurrently by collaborative partners, without any conflicts, and that security issues discussed earlier are not sacrificed.

Redundancy. Permanent data loss in the event of an incident must be avoided. The architecture must include redundancy measures to maintain backups of data models and user analysis results.

Client footprint. It should be assumed that FACT will be accessed largely from DoD client computers and that all software components will need to comply with inherent restrictions. Typical government users are unable to install software to their machines, therefore FACT must require no client software installation in order to be used, and should utilize only those software components which are installed by default on specific government computers.

2.3 Architecture Requirements

Software architecture development must start with proper requirements development, which ultimately supports test and verification. Initial requirements for FACT will be outlined, and should be further developed and solidified as this effort progresses. Although specific requirements are not yet defined, a software architecture meeting the demands defined in this paper should be capable of the following notional requirements, to include being able to:

- provide a near real-time user experience for M&S-based cost and performance analysis.
- be equally accessible to all identified user bases (i.e. government, academia, and industry).
- support many concurrent user sessions without degradation to an unacceptable level of performance.
- retain users' analysis results, and enable future retrieval.
- use a model-independent system architecture to support future hosting and analysis of an arbitrary number of models.
- include an authorization system allowing establishment of user roles with varying permissions levels.
- provide interfaces and processes for the inclusion of federated data, and provide an acceptable level of performance for its analysis.
- have data storage redundancy at multiple offsite locations.
- be a true "zero-footprint client" web application, meaning no client-side installation beyond common software
- be built using a modular system architecture, to facilitate scalability, expansion, and future modification.
- provide the capability for users to run high-performance computing processes.

2.3.1 Server Architecture

An initial vision for providing a web interface-based, rapid performance and cost trade space analysis is discussed in this section. The physical architecture provided

in Figure 1 is introduced to foster discussion for the eventual FACT functional and physical architecture. The proposed physical architecture follows the standard three-tier framework common among web services. At the top of the stack is the client (which may be any number of instantaneous clients) that requires the following components: a security authorization device, a compatible Internet browser, and may or may not have desktop applications including an authoring tool for defining and manipulating actual vehicle architecture considerations.

The selected Internet browser must support client scripting, specifically JavaScript, for the generation of the user interface and handling asynchronous data passing between the server and client. Scripting solutions for both two- and three-dimensional visualizations are planned. Each client is connected to the web server via Hypertext Transfer Protocol Secure (HTTPS) through a firewall to provide encrypted communication and secure identification of a network web server. The web server houses the Hypertext Transfer Protocol Daemon (HTTPD) server, an interpreter for server scripting language for computation and data transfer between the database server, a queue server for asynchronous message passing, and version controlling software. Considering that the problems of interest are complex vehicles, a server scripting language that provides or supports numerical and symbolic processing is necessary.

Finally, the web server communicates with the database server (protected by a second firewall) via the Structured Query Language (SQL) or the SPARQL Protocol and Resource Description Framework Query Language (SPARQL) queries. The database server will, at a minimum, house a relational database management systems (RDBMS) and may house a semantic data store intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.

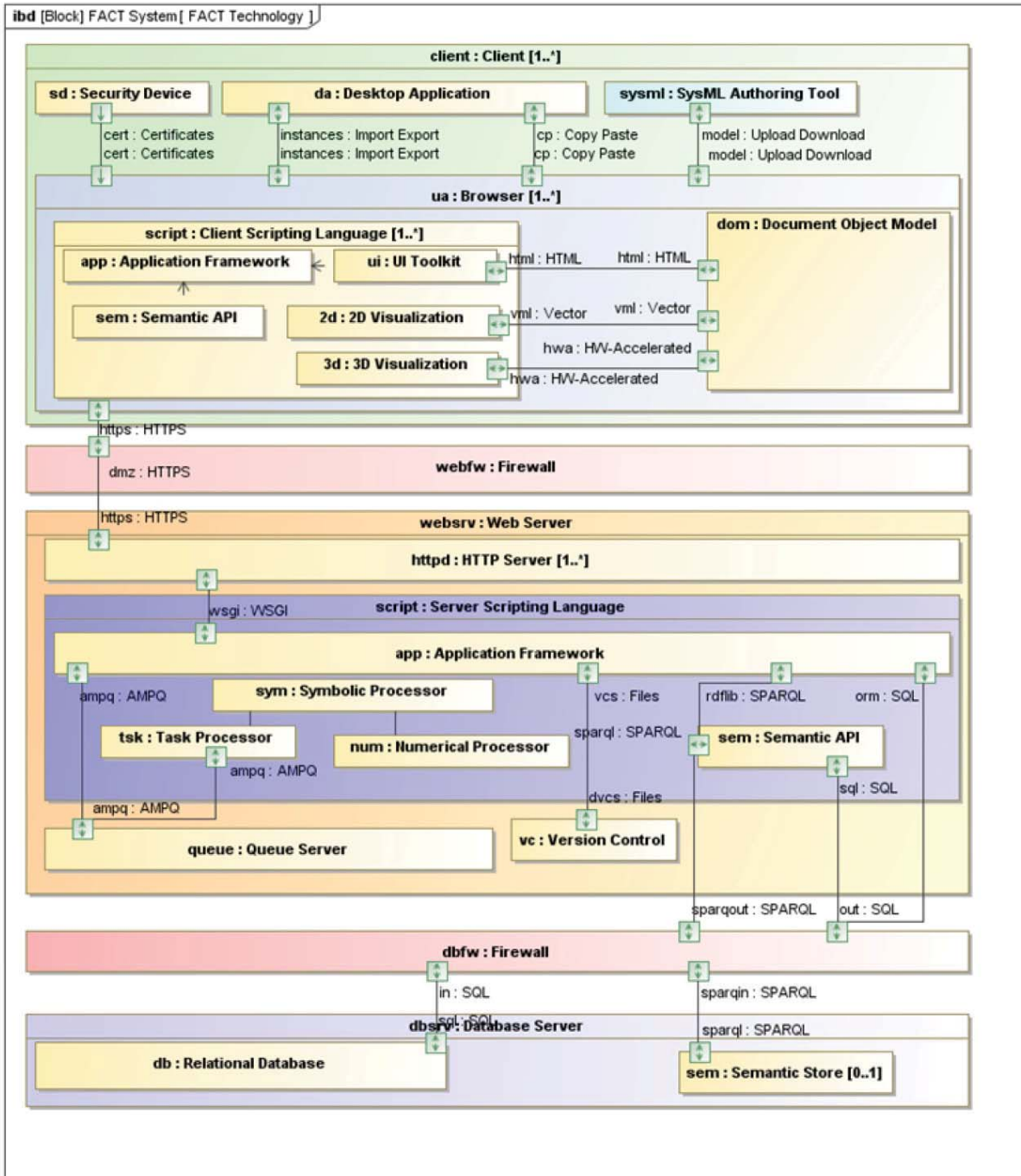


Figure 1. Proposed Server Architecture.

2.3.2 Software Architecture

A core goal for the FACT is to be a center for M&S-based cost and performance analysis, for an arbitrary number of models. Interfaces must be open and common, and a modular approach to the architecture will enable easy expansion and additional

capability in the future. The use of open-source software tools is in alignment with this approach, as discussed earlier, and is also consistent with the DoD's increased emphasis on reducing total ownership cost. As described above, the proposed approach will follow the common web services

paradigm of a database layer, a web application layer, and a presentation/visualization layer. Given that the architecture will use a modular design, all of the proposed software packages are free and open-source; the specific software architecture will remain fluid during this exploratory phase. Ultimately, various options will be explored and tested to find the configuration with optimal performance.

Database Layer. Relational databases are the most common way to store data; however, semantic stores, often used to enhance web search capability, are becoming more popular as they provide better constructs for describing data. The FACT will need to maintain a data store for each of these types of databases. These databases will reside on their own node, separate from the rest of the architecture. Hardware infrastructure will be constructed to enable easy addition of capability to this node (or to expand onto multiple nodes). Open-source options exist for RDBMS, such as MySQL, which are free, scalable, and mature in development and capability to meet the requirements defined in this paper.

Web Services and Computational Layer. This portion of the system architecture will reside on its own node. It will connect to the database server via an internal network, and will include a web server that will be visible to outside connections. Like the database server, infrastructure will enable easy expansion in the future if more resources are required. Regarding the computational layer, a Python-based open-source stack has strong support for those computations typically required by an M&S application (i.e. numerical analysis and linear algebra). In addition, many free Python libraries exist for direct interaction with the semantic and relational databases that will be located on the FACT database server.

To provide in-house support of complex M&S execution, high-performance computational nodes fitted with general purpose graphics processing units (GPGPU) is envisioned. These nodes can provide teraflop performance at lower cost

than general purpose processors and allow for inclusion of models that cannot be abstracted to surrogate representations.

Client Presentation/Visualization Layer. The presentation layer contains the web content that will be displayed to the user. In order to enhance the user experience, a rich internet application (RIA) framework will be used to develop the actual code needed to create the web content. A number of RIA framework options exist for creating rich web content. The preferable option is Open Standard, in keeping with the rest of the FACT architecture. Usage of closed-standard frameworks like Adobe Flex and Microsoft Silverlight is widespread, but those would impose various limitations. However, those closed standard options may be utilized for parts of the GUI if specific visualization effects, such as 3D renderings, are desired.

2.4 Integration with M&S

The long term goal of the FACT effort is to provide a full lifecycle analytical capability to support program success. The FACT process will enable integration with a suite of M&S tools, allowing decision makers to compare and prioritize technology, generate valid design decisions, conduct design space exploration and compare valid design solutions. However, a key element of the approach put forth in this paper is the ability to conduct “rapid” analysis. Conducting M&S of complex systems may be time consuming, often limiting the amount of data that can be generated in a practical amount of time to warrant a usable trade study. This is most certainly the case for high-fidelity, federated simulations.

The primary enabler of rapid, dynamic integration of any decision support toolset with M&S is through the use of surrogate modeling [9]. Surrogate models are equation regression model representations developed through sampling of more complex computer based M&S tools. Sampling is guided by selection of an appropriate Design of Experiments (DoE). Surrogate modeling has been

applied in the development of M&S-based decision support tools by one of the authors, most recently documented in support of ballistic missile defense architecture analysis [10]. Early adoption by industry and government has proven successful in developing fuel efficient military ground vehicles [11].

A component of the FACT process is a set of visualization tools that show the

inter-relationships of design parameters, cost, and reliability, shown notionally in Figure 2. Leveraging surrogate modeling will enable a display of the inter-relationship of design and resource parameters, and will present a sensitivity analysis of how changes in those parameters affect the overall cost, reliability and performance of a system.

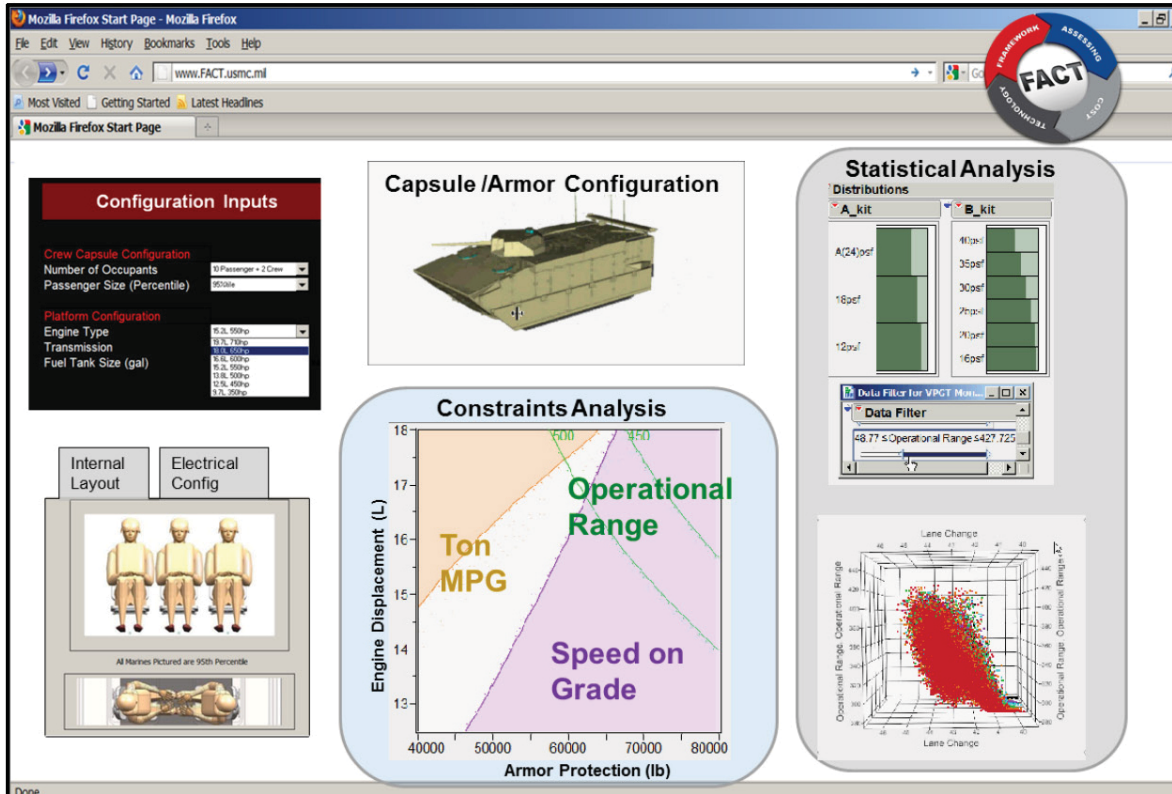


Figure 2. Notional web browser-based FACT “dashboard”.

3.0 CONCLUSION

Decision makers are often forced to trade requirements with limited information; however, the only way to truly understand the trade-off between requirements is to understand the physics of the problem. Acquisition of complex systems has therefore made modeling and simulation a necessary element of the development process. With a number of notable acquisition program cancellations, the DoD is moving to better inform decision makers and program managers.

This paper introduced the Framework for Assessing Cost and Technology: a vision for enabling enterprise-level analysis leveraging the knowledge gained using modeling and simulation, and collaborative nature of web-based tools accessing databases hosted on secure servers. Guidelines for this effort were discussed and shown to support the broader context of fostering Community of Interest-wide discoverability, coordination, interoperability, and reuse of data. Software architecture considerations and requirements were discussed, and details related to web-based tools and server

architecture was reviewed. An emphasis was placed on open-source software, given recent acceptance by the DoD community for adaptability and viability as a cost savings measure.

Attention was given to the use of surrogate modeling as means to capture and directly, and dynamically, integrate the effects of M&S within broader decision making. This enables rapid analysis of alternative technology/materiel through the use of surrogate models, or equation regression representations of more complex M&S tools, as proven through several successful implementations discussed in this paper.

The use of FACT during the materiel solution and analysis phases of development will help to develop a set of "informed" requirements. These "informed" requirements will prevent a program office from embarking on a design path that will be high risk or unattainable based on the maturity of the technology proposed for the system. This process will ultimately reduce program development and life cycle costs, both of which are tenets of effective "should-cost" management.

4.0 REFERENCES

- [1] United States Congress, Section 222 of "National Defense Authorization Act for Fiscal Year 2010," Public Law 111-84, Washington, DC, 2010.
- [2] Under Secretary of Defense, Acquisition, Technology, and Logistics, "Memorandum on Implementation of Will-Cost and Should-Cost Management," 2011.DoD Modeling and Simulation Coordination Office, "Modeling and Simulation Community of Interest Discovery Metadata Specification (MSC-DMS)," Version 1.3, 2010.
- [3] DoD Chief Information Officer Memorandum, "DoD Net-Centric Data Strategy," 2003.
- [4] DoD Directive 8320.02, "Data Sharing in a Net-Centric Department of Defense," 2007.

- [5] Office of the Director of Defense, Research, and Engineering, "Strategic Vision for DoD Modeling and Simulation," 2007.
- [6] DoD Chief Information Officer Memorandum, "Clarifying Guidance Regarding Open Source Software," 2009.
- [7] Military Open Source Software, homepage: <http://mil-oss.org>.
- [8] Assistant Secretary of Defense (Networks and Information Integration) and DoD Chief Information Officer, "DoD Information Assurance Certification and Accreditation Process (DIACAP)," Department of Defense Instruction 8510.01, November 28, 2007.
- [9] Forrester, A., Sóbester, A. and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*. West Sussex, John Wiley & Sons, Inc., 2008.
- [10] Ender, T., Leurck, R., Weaver, B., Miceli, P., Blair, W.D., West, P., Mavris, D., "Systems-of-systems analysis of ballistic missile defense architecture effectiveness through surrogate modeling and Simulation," *IEEE Systems Journal*, vol. 4 no.2, pp. 156-166, June 2010.
- [11] Luskin, P. and Berlin, R., "Systems Engineering Methodology for Fuel Efficiency and its Application to the TARDEC Fuel Efficient Demonstrator (FED) Program," *Proceedings of the 2010 NDIA Ground Vehicle Systems Engineering and Technology Symposium*, Dearborn, MI, 17-19 August, 2010.

5.0 ACKNOWLEDGMENTS

The authors wish to thank the Marine Corps Systems Command for funding this effort. The authors also wish to thank Ms. Lorri Bennet and Dr. Jay Martin of the Pennsylvania State University Applied Research Laboratory for lifecycle cost insight, and related guidance on architecture considerations. Finally, the authors wish to thank Ms. Jennifer DeLockery, Mr. Drew Pihera, and Mr. Jim Bertoglio of the Georgia Tech Research Institute for development of the FACT architecture implementation represented in this paper, and for crafting the overall vision shaping this effort.

1.15 A First Person Shooter/Real Time Strategy Hybrid: Lessons Learned

A First Person Shooter/Real Time Strategy Hybrid: Lessons Learned

James R. Stoup, Olen Bruce

JTCOIC SIMS

Newport News, VA

jstoup@quantum-intl.com,

obruce@alionscience.com

Bobby Jones

JTCOIC SIMS

Newport News, VA

bjones@alionscience.com

INTRODUCTION

Today's military training bears little resemblance to the methods of previous generations. The Cold War is over and the enemy has changed. Doctrine that was once useful is now woefully out of date. No longer are we confronting a predictable nation but instead a diverse collection of independent fighters spread out over several countries. The enemy has changed, his tactics have changed and circumstance dictates that we must change as well.

The wars in Iraq and Afghanistan have tested virtually all aspects of the military's support infrastructure. After fighting continuously for over a decade, many weaknesses have been revealed by the steady grind of war. Chief among them is the inability of the military to rapidly and adequately train its soldiers in the latest doctrines.

In response to the enemy developing new strategies on a near monthly basis, the Joint Training Counter-IED Operations Integration Center (JTCOIC) was first created. Designed to supplement the current training system, it would attempt to address the current training shortfalls. As a result, new methods were devised to streamline training.

THE TRAINING CHALLENGE

Currently the military's training program is in flux. It is in the process of transitioning away from older table-top methods and towards a completely new digital paradigm, a move made possible only recently by the widespread adoption of new simulations technology. However, to fully appreciate the significance of this change it is helpful to look at how the military trains today and how it hopes to train tomorrow.

Current U.S. Military Training Doctrine

Currently the most prevalent form of training is classroom lecture supplemented with PowerPoint presentations. This method is popular because it is

cheap, simple and easily customizable. Additionally it can be designed to accommodate almost any size class and requires virtually no additional materials other than a computer and projector. Unfortunately, this method can also be fairly inefficient (Worley & Dyrud 2004). Instructors are limited by their experience, the curriculum is often out of date and, most importantly, students are easily fatigued by long sessions. Often the results are students, overwhelmed by data, who ultimately retain very little from the experience (Tufte 2003). So while this method is cheap and easy, it is also does a poor job effectively training soldiers.

In support of stateside training, and to supplement lecture based instruction, Mobile Training Teams (MTT) are sent into the field to train soldiers in theater. These teams do an excellent job; however they are tasked with training an incredibly large number of soldiers, in different countries, in wildly differing environments. While this is preferable to the alternative of no training, providing a consistent training experience for all troops has proven to be quite difficult. Despite an increase in manning, there simply isn't enough MTT's to fulfill the demand.

Recognizing that the increasing frequency and diversity of U.S. military conflicts will continue to require highly trained soldiers, steps have been taken to modernize the current training environment. In keeping with this realization, a new directive was announced in 2010 aimed at formally expressing this long awaited change in direction.

The U.S. Army Learning Concept for 2015

In a letter explaining this new training concept for 2015, Training and Doctrine Command (TRADOC) commander General Dempsey described it as follows:

We live in a much more competitive security environment. This means that we shall have to learn faster and better than our future adversaries. Stated a bit differently, we must prevail in the competitive learning environment.

...
The Army Learning Concept 2015 does not focus on any particular technology, but rather focuses on the opportunities presented by dynamic virtual environments, by on-line gaming, and by mobile learning. It speaks of access to applications, the blending of physical and virtual collaborative environments, and learning outcomes.

*Martin E. Dempsey
General, United States Army
Commanding*

This letter was written to clarify the purpose of this new doctrine. Specifically, it mentions how the military plans on improving every aspect of their soldier's training. New tools will be used that places a premium on mobility, virtualization and universal data access.

Since this new paradigm of military training focuses on the key aspects of adaptability, mobility and up to date material, the Maneuver Center of Excellence (MCOE) team decided to base part of their training curriculum on the tactical simulation, Virtual BattleSpace 2 (VBS2 2011).

THE MCOE EXPERIMENT

Computer based training simulations offer a rapid and efficient way to train a large number of soldiers. Moreover the current soldier is quite familiar with this type of training as they have played commercial video games for many years, making them particularly well suited for this medium. As such, this type of training is gaining attention in schoolhouses looking to modernize their curriculum (Sellers 2001).

In an effort at revamping their curriculum, the Maneuver Center of Excellence (MCOE) decided to rework the Leadership Decision Exercise portion of their training program. Specifically, they began laying plans to incorporate Virtual Battle Space 2 (VBS2) into their classes. To assist with that effort, MCOE instructors met with the Systems Integration Modeling & Simulation (SIMS) division of the JTCOIC.

MCOE instructors would meet many times with JTCOIC SIMS programmers as they crafted a series of requirements that they would need VBS2 to fulfill if the training was to be a success. Presented below are the key elements they agreed upon:

- Capability to train multiple soldiers at once
- Capability to train multiple command positions simultaneously
- Ability to play from enemy perspective
- Realistic map based interface
- Use of geo-specific terrain
- Large scale training with minimal staffing
- AAR capability

These objectives align with the Army Learning Concept 2015 parameters and taken together, these objectives outline a method of training soldiers that will be far superior to the decades old sandbox method currently in place.

The solution that was ultimately delivered to the MCOE consisted of a highly customized training scenario run in VBS2. The scenario allowed for over 11 students controlling over 200 entities to participate in a coordinated offensive against an enemy building in an urban environment.

To achieve this coordination, the standard First Person Shooter (FPS) interface of VBS2 was enhanced by adding in the map based control structure of a Real Time Strategy (RTS) game. Before we can explain the details of this solution though, an explanation of the similarities and differences between a First Person Shooter and a Real Time Strategy game is required.

First Person Shooters

Because of the popularity of such commercial FPS titles like *Half Life*, *Halo*, *Counter Strike* or *Medal of Honor*, this genre of game is quite familiar to this current generation of soldiers. Though this particular style of shooter has been around for over 20 years, it wasn't adopted by the Military for training simulation until fairly recently. The current best selling FPS (see Figure 1) is shown below.



**Figure 1. Most Popular Current Commercial FPS,
*Call of Duty 4***

The current standard FPS training simulator in the military is *Virtual Battle Space 2* (VBS2), a military version of the commercial game, *Armed Assault*. This simulator (see Figure 2) has proven to be so useful that it is currently used by all four branches of the military in addition to numerous other coalition partners such as the U.K., Canada, New Zealand and Australia.



Figure 2. U.S. Army FPS of Choice, *VBS2*

VBS2 is primarily used for squad level training. Designed to support roughly 200 in game units at once on a map 10 km square, the scope of the game is limited to a city sized game area.

Because the perspective of the player is confined to one unit, to complete a mission multiple human players are required to play simultaneously via a network game. AI units are used to supplement the experience; nonetheless there is a direct correlation between complexity of a scenario and the number of human players required.

Real Time Strategy

Unlike First Person Shooters which focus solely on the individual, Real Time Strategy games (see Figure 3) focus on controlling large armies simultaneously. There is a diverse mix of units, a large terrain to traverse and effective play requires coordinating multiple combined arms attack groups.



Figure 3. *Company of Heroes*

RTS games are the commercial equivalent of the military's virtual constructive simulations such as JSAF or JCATS. With a focus on coordination of forces and choice of tactics, these simulations are better suited to training commanding officers.



Figure 4. *Command & Conquer 3*

The two examples shown above depict two very popular RTS style games, *Company of Heroes* and *Command & Conquer 3*. In each game there is a map (see Figure 4) used to direct a player's forces along with various other methods for controlling units. This map based interface would prove to be the critical component of the MCOE solution.

CHALLENGES FACED

Now that the framework has been laid for understanding the MCOE training scenario it is time to

move on to analyzing the problems this team encountered, along with the key lessons they learned as this development proceeded.

There were four main areas of expertise that came together to create this project. Each area tapped the resources of different teams within the JTCOIC SIMS. As a result, even two members of SIMS completed the bulk of the work, this project would eventually involve contributions from almost a dozen SIMS personnel. This project would come to include scenario developers, programmers, 3D artist, terrain specialists and system administrators as everyone worked feverishly to meet an April 14th deadline.

The remaining portion of this document describes the lessons learned by each team as they struggled to complete a high quality product in the face of changing requirements, increasing project scope and unforeseeable difficulties.

LESSONS LEARNED - OPERATIONS

Before any code was written, before any graphics were designed or any network setup, the entire training exercise had to first be designed. Fort Benning's MCOE leadership would work closely for over nine months with JTCOIC SIMS personnel as the MCOE training scenario was developed.

Have a Process to Determine Requirements

The first lesson learned also turned out to be the most important, namely determining project requirements. In what would become a long and detail oriented process the MCOE instructors slowly worked out what they wanted this training to entail.

What they ultimately decided on was a VBS2 Urban Ops scenario set in downtown Columbus, Georgia. Enemy combatants had taken over a walled public safety building complex and four platoons of infantry and coalition forces had been tasked with securing the area. The students would play various command roles within the BLUE forces such as company commander, aviation controller, platoon leaders, squad leaders or OPFOR. The ultimate goal was to retake the prison, kill or capture all enemy units and do so with minimal loss of life and restore control to the local government authorities.

This mission template was initially developed over a period of several days. However, senior leaders within MCOE and other CoE's such as Fort Rucker and Fort Sill expressed a desire to participate in the training

which resulted in new requirements being added. Additional changes were made as they discovered what the simulation could fully do. As a result, instead of being given a single document containing the mission specifications, what was produced was a constantly evolving set of requirements. This more than anything contributed to the length of time needed to produce a working version.

Understand the Depth of the Problem

Further delay was caused by the training staff underestimating the amount of operational data the programmers would need. Because the development team had no prior military experience to use as a reference point, instructions that were self explanatory to military personnel required elaboration for the programmers.

For instance, it was insufficient to merely state that a unit had to move to a new location. Additional details were needed such as convoy movement patterns, alternate destinations if initial one was blocked, call signs of the units, correct terminology for movement orders and verbal responses.

Often times seemingly simple mission features could require much more supplemental information than initially provided. Data such as armaments for the air support, accurate map icons or proper breaching procedures were some of many subtle details that had to be right for the training to be accurate.

Solutions

For future training scenarios, explicit objectives must be defined before any other work can begin. Moreover, it is required that someone with advanced VBS2 experience be involved with the initial planning to provide a reasonable guide for what the simulation is and is not capable of producing.

Then, once the initial framework of the scenario has been complete, as many mission details as possible should be recorded. Attack plans need to be diagramed, convoy routes planned, medevac and air support defined along with any custom details that would be required. By providing as much information as possible up front, mission designers can minimize the need to rework the scenario later in development.

LESSONS LEARNED - TERRAIN

Building a geo-specific terrain (a terrain modeled on an actual location) of Columbus, Georgia was a long and

challenging process made harder by the lack of American building models in VBS2. Because VBS2 is used extensively to train soldiers going to Iraq and Afghanistan most of the terrain features and buildings are specific to Iraq and Afghanistan. The landscape leans towards deserts, mountains and minimal greenery while the buildings are squat, drab and brown. Though this has worked quite well for its intended purpose, it doesn't blend very well in urban America.

As a result, custom terrain had to be designed pulling in as many geo-typical American looking buildings, foliage and unit models as could be found, some of which had to be created from scratch.

Use Location Specific Models

Due to a lack of American looking houses and buildings, many Eastern European models had to be substituted. Some buildings such as the prison complex were designed in house for use by the terrain team. And while this solution was satisfactory, if future training missions were to be conducted in American cities, it would be worthwhile to develop American building, car and tree models. This would make the terrains more realistic and provide greater visual variety.

Smaller Maps are Better

Due to the limits of VBS2, the largest terrain it can currently handle is 50k x 50k. That size terrain is undesirable though because it wouldn't have sufficient detail to accurately represent the training area and would be a tremendous drain on system resources. Instead, smaller, denser maps are preferred as they provide better training experiences.

For this mission a 10k x 10k map was used. The main combat area was a 0.5 square mile section (which accounted for about 9 city blocks) that was created at a higher resolution than the remaining city. This increase in density made for a more realistic setting by including more models such as cars, trees, signposts, light fixtures and debris. Additionally the buildings in this section are of a better quality than those outside the combat zone. And because the majority of combat operations occurred within this zone, improving the fidelity of this area greatly increased the quality of the simulation.

Of course, even at this much reduced map size details begin to be lost the further away from the city center. This terrain required 2 man-months of work to complete. The chart below represents various terrain sizes. In determining how large a terrain is needed, it

is useful to compare the size of terrain (see Figure 5) that was created for MCOE (10km, green box) with the area of the city that was actually used (.5 miles, black box).

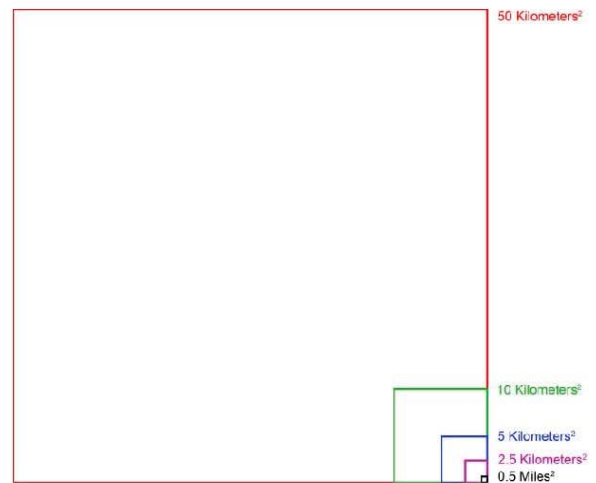


Figure 5. Relative Terrain Sizes, VBS2

Solutions

Taking the time to model American style buildings, roads, cars, trees and other misc. bits of scenery will greatly increase the quality of the terrains. Additionally, adding these types of models is recommended as it is expected that more missions like this will be created.

Focusing on smaller, denser maps will allow terrain designers to make more realistic terrains in less time. Smaller maps also make better use of the limited number of polygons allowing for more realistic environments.

LESSONS LEARNED - SCRIPTING

Once the requirements were completed and the terrain built, the next step was to write the code that would provide the needed functionality. There were two main components needed to control the designated units, the Graphical User Interface (GUI) code and the unit control code.

The GUI code centered on creating a map that would have elements that would allow for the controlling of troops and vehicles.

The control code took the inputs from the map and translated them into commands that would move the

Multiplayer Issues Change Everything

Correct Path Planning

Complicating matters are situations involving multiple units trying to arrive at the same point simultaneously. Some type of stacking algorithm must be employed to prevent a backup of vehicles.

Creating the MCOE Control Map GUI (see Figure 7) was a significant challenge. Developing a system to provide the desired functionality required considerable development time. Further complications ensued when modifications were needed of the map itself along with the list of units to move and actions to perform.

Solutions

In developing the map interface it was discovered that using an open architecture is preferred. Easily extensible maps offer the user better control and can be developed faster. Specifically the map should be dynamic, not have fixed points (the blue dots in the map image) that the units can move to.

Once development was complete it was time to field the system for testing by the customer. The scenario would run on a set of 14 student laptops connected to a server hosting the scenario. And even though the software was complete, there were still several problems that would have to be solved for the mission to proceed.

One of the most important lessons learned is that a mission PBO file loads much faster than an editable scenario. PBO files remain the native binary file

format of VBS2 content, while an editable mission is instead a collection of files in the %userprofile%\VBS2\mpmissions directory.

When a networked scenario is invoked, the scenario server copies all required files to the connected clients. The MCOE scenario contained a significant number of custom files such as images, audio files and scripts. Unfortunately, VBS2 allocates a maximum bandwidth of approximately 1.5Mb/sec for the server to copy editable scenario files. This maximum bandwidth is shared between all participating clients. Thus the time required to transmit a copy of the mission to each computer depends mainly on the size of the files and the number of networked computers. The function defining this relationship is given below.

The scenario load time in minutes (L_m) is given by:

$$L_m = \frac{[(S_z * 8) * N_c]}{(60 * B_d)}$$

Where S_z is scenario size in MB, N_c is number of connected clients, and B_d is the VBS2 maximum bandwidth in Mb/sec.

Calculating L_m for the MCOE scenario produced the following results:

$$24.57 = \frac{[(15.8 * 8) * 14]}{(60 * 1.2)}$$

To transmit the scenario to every computer took almost 25 minutes, this is an unacceptable delay. The bottleneck here neither the operating systems nor the physical network, it is the maximum bandwidth VBS2 allocates for copying scenarios. Using any other method, copying a 15.8 MB VBS2 mission scenario to 14 clients on a 1Gb/sec network takes less than 60 seconds.

To avoid the slowdown associated with transmitting mission files, the mission should be saved as a PBO and then manually copied to each machine. Only Mission PBOs can be read locally, copying the Mission directories will result in the server pushing out the files anyway, causing the 25 min. delay.

Throughput is Everything

A poor network setup caused the game to become unusable. The setup as it was initially designed is shown below (see Figure 8).

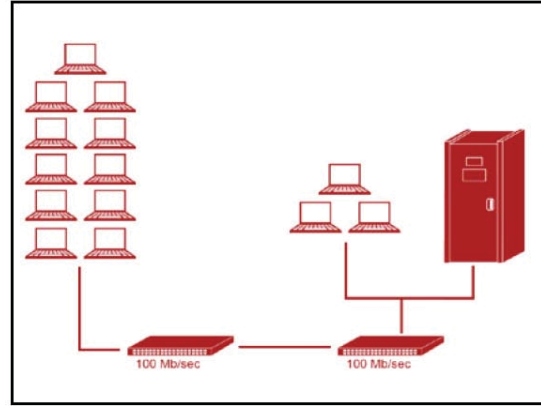


Figure 8. Initial Computer Setup

Because the majority of traffic was going through two slower switches, the server couldn't keep up with the traffic and system crippling lag was the result. The better setup is shown below (see Figure 9). It has everything streamlined to prevent any of the bottlenecks that affected the original attempt and a faster switch is used to ensure sufficient bandwidth.

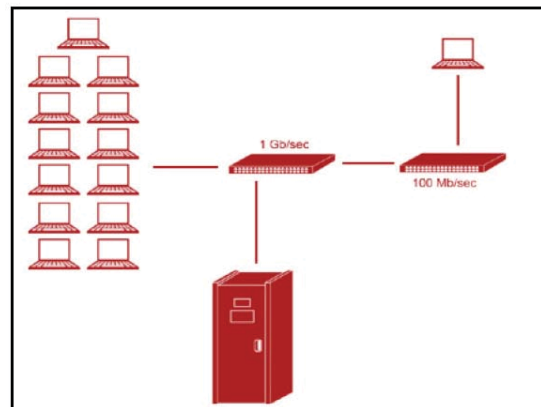


Figure 9. Final Computer Setup

Networking Other Applications

While VBS2 was the main simulation application in use, there were other applications used for specific tasks supporting the MCOE simulation scenario.

Ventrilo was used for VoIP (Voice over IP) communications. The server component was loaded on the VBS2 server while all client computers had the client application. According to the Ventrilo server requirements, each client (voice stream) could consume 8KB/sec (112KB/sec for fourteen clients).

The Situational Awareness Translator (SAT) is used to provide an LVC (Live, Virtual, and Constructive)

mechanism via DIS (Distributed Interactive Simulation) broadcast packets which allowed the correlated translation of entities between VBS2 (running in LVC mode) and FBCB2 (Force XXI Battle Command Brigade and Below).

Solutions

The greatest impact on performance was the network the machines ran on. Since the server and all laptops were running on new hardware, the slowdowns could only be attributed to bottlenecks in the network. Once the network was properly configured and using switches with sufficient bandwidth, the problems disappeared.

CONCLUSIONS

After nine months of development a final version of the MCOE training scenario was delivered. Though there were many challenges along the way, the core simulation that was originally planned was successfully deployed. Moreover, once integrated into the training curriculum it provided a high quality training experience that was met with approval from both the instructors and the students.

MCOE can be viewed as a pilot program designed to gauge the feasibility of improving specific types of training using VBS2. Because of the success of this project, more Centers of Excellence have agreed to participate in the next iteration of this training.

However, any organization that would consider replicating this training would be well advised to make use of the lessons learned by the JTCOIC SIMS.

- It is vital that all requirements be generated early on, with specific training goals defined.
- Once these requirements have been determined and agreed to, highly detailed battle plans, maps, protocols and mission parameters must be provided.
- Only then can the mission be programmed and the terrain crafted. After extensive testing with the initial planners, the simulation must be correctly compiled and distributed to all required machines.

- A full dress rehearsal with instructors is a requirement. All key leaders must know the capabilities and limitations of the scenario.
- Only after ensuring that the network has been properly configured can the training scenario start.

In conclusion, using these lessons learned should provide a dramatic shortcut to anyone looking to expand this type of training.

ACKNOWLEDGEMENTS

The MCOE training product was the result of many hours of hard work by the entire JTCOIC SIMS team. Specific thanks go to Colonel Zeisman, Major Curry, Darien Hawkins, Larry Reynolds, Valerie Plyler, Jonah Atkins, James Kadra, Zach Phillips, Mike Barrington, Jeff Bittel and Mark Covey.

REFERENCES

- Dempsey, Martin E., General, United States Army Commanding (2011). Personal Communication.
- Worley, R. B., & Dyrud, M. A. (2004). Presentations and the powerpoint problem. *Business Communication Quarterly*, 67(1), 78-80.
- Tufte, Edward (2003). PowerPoint is Evil. *Wired Magazine*. Retrieved from <http://www.wired.com/wired/archive/11.09/ppt2.html>
- Sellers, Russel. (2011). Army Learning Concept 2015 brings Aviation, ground forces together early. *U.S. Army News*. Retrieved from <http://www.army.mil/article/54953/>

1.16 Framework for Architecture Trade Study using MBSE and Performance Simulation

Framework for Architecture Trade Study using MBSE and Performance Simulation

Jessica Ryan
George Washington University
Jryan82@gwmail.gwu.edu

Dr. Shahram Sarkani, Ph.D., P.E.
Professor of Engineering
Management and Systems
Engineering

Dr. Thomas Mazzuchini D.Sc.
Professor of Engineering
Management and Systems
Engineering and Decision Science

Abstract. Increasing complexity in modern systems as well as cost and schedule constraints require a new paradigm of system engineering to fulfill stakeholder needs. Challenges facing efficient trade studies include poor tool interoperability, lack of simulation coordination (design parameters) and requirements flowdown. A recent trend toward Model Based System Engineering (MBSE) includes flexible architecture definition, program documentation, requirements traceability and system engineering reuse. As a new domain MBSE still lacks governing standards and commonly accepted frameworks. This paper proposes a framework for efficient architecture definition using MBSE in conjunction with Domain Specific simulation to evaluate trade studies. A general framework is provided followed with a specific example including a method for designing a trade study, defining candidate architectures, planning simulations to fulfill requirements and finally a weighted decision analysis to optimize system objectives.

1 NOMENCLATURE

2 INTRODUCTION

The foundation of MBSE is an object oriented design process which uses heterogeneous modeling techniques to capture system architecture, relationships, requirements and constraints [1, 2]. In many aspects MBSE complements the classical System Engineering approaches, (i.e. Waterfall, Standard 'Vee', Spiral), however document-centric processes are replaced with models which offer traceability, various viewpoints and a central repository for design information [3]. By taking advantage of the object oriented structure of MBSE, and improvements in simulation tool interoperability, a novel framework is proposed which optimizes Architecture Trade Study through MBSE and Performance Simulation integration.

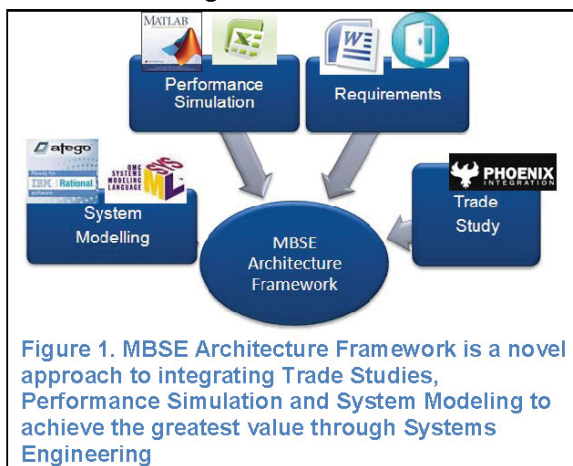
Section 2 is a literature and concept review of the influencing factors in this framework.

Section 3 provides a framework and demonstrates a potential application of the framework using the available tools, processes, and methods common to MBSE. A trade study is presented which varies the design parameters of a data recording system given requirements associated with cost and performance. The fundamental architecture is composed of a receiver, an analog to digital converter (A/D), a high speed buffer and a solid state data device (SSR). Section 4 is a discussion of the results, showing the benefits of the Architecture Framework using MBSE and offering a path forward for additional research.

2.1 MBSE

As defined by INCOSE "Systems Engineering is an interdisciplinary approach and means to enable the realization of successful systems" [4]. MBSE attempts to optimize the design, implementation, delivery and operation of a system throughout its entire lifecycle through modeling techniques as opposed to a standard document-centric approach.

Parameters of the system that provide the maximum benefit vary depending on the actual stakeholder and can range from reduction of cost, reduction of schedule,



increased performance, sustainability, reliability, etc [5]. As each stakeholder (customer, end user, designer, etc..) has different viewpoints in relation to the system, multiple diagrams and associations exist to address system aspects [6, 7]. Modelling of interfaces at multiple levels of abstraction aids understanding system complexity. Underlying links ensure communication of the current design which is important for configuration control, insight into design assumptions, and justification of requirements allocation.. Success from the software engineering domain using UML motivated the profile extension to SysML which now includes all of UML as well as a Requirement Block Diagram and Parametric Block Diagram [8, 9]. SysML is an extension of UML based on the following four fundamental pillars: Structure, Behavior, Requirements, Parametrics. The example presented in this paper is designed using Artisan Studio SysML Profile, however this design decision is independent of the fundamental principles of the framework. The scope of this paper is limited to the usage of this language, for further description of the semantics and definitions refer to [10].

2.2 Domain Specific Simulation

Domain specific simulations are used to evaluate behavior of the system for a given set of inputs, typically requiring custom tools which can execute various system responses (continuous or non-continuous). While SysML does offer features such as automatic code generation and requirement traceability automation, SysML is not an executable language per se. The purpose of SysML is to describe the system (including architecture, requirements, associations, driving function) but is not necessarily designed to execute or evaluate simulations tied to the system behavioral response. However, because of its meta-language base, SysML can be extended to interact with domain specific tools [6, 11]. Tool interoperability remains a great challenge when modeling system parameters. Most standard tools sets (ie Matlab, Excel, Agilent ADS) offer APIs which allow various levels of interaction with the models. Co-

simulation between SysML and common industry performance analysis tools such as Simulink and Modelica is an active topic of research [12, 13]. Co-simulation can be executed in 2 manners : 1) code generation from SysML, post simulation with Domain Specific code 2) graph transformation [14, 15]. In either case, supporting custom code to bridge the gap between Domain Specific tools and SysML is required.

2.2.1 Analysis of Alternatives

Trade studies are used during an Analysis of Alternatives in a manner to determine the best system architecture [16]. Modeling and simulation are often used during an Analysis of Alternative (AoA) as a cost-effective means to assess design trades and understand their impact on system response. A reform in general defense acquisition has led to a greater appreciation for system maturity which can be rated as a System Readiness Level [17]. This paradigm shift from Performance-Based Acquisition to Capabilities-Based Acquisition shows the complexity in requirements analysis [18]. AoA can be used to assess architectures representing various System Readiness Levels factors thereby supporting Capabilities Based Acquisition.

3 FRAMEWORK

This framework provides a methodology which uses MBSE to guide the structure of Performance Simulation and the efficiency of Trade Studies to define System Architecture. To date, several framework exists for MBSE including IBM Telelogic Harmony-SE, INCOSE Object-Oriented Systems Engineering Method (OOSEM), IBM Rational Unified Process for Systems Engineering (RUP SE), Vitech Model-Based System Engineering (MBSE) Methodolog, and JPL State Analysis (SA), yet there is no uniformly accepted standard [3]. While literature exists to support integration of system modelling through SysML to domain specific tools, a framework does not yet exist to provide utility

in architecture trade study using analysis of several models.

Trade Studies are improved through identifying the driving input parameters and desired system requirements. Requirements analysis metrics are used to assess each trade response and verify the ability to meet the traced requirements.

Performance simulation benefits from abstraction as parameters including inputs, outputs, and driving interdependencies (i.e. number of bits <nBits> for an A/D) are identified for each block (hardware component). In an object-oriented fashion, these parameters are abstracted such that any A/D would have a given set of parameters. A particular instance of an A/D has a discrete value for each parameter (relationship between a class versus instance). After inputs, output, interdependencies, and purposes for each simulation are identified, various algorithms can be tested. Defining simulation interfaces gives flexibility to later increase the fidelity of the model, or change methods by which the outputs are determined (ie various methods of SNR calculation).

3.1 Process

The framework follows a process which emphasizes requirements analysis, trade study definition, and performance simulation to determine the optimal architecture. The following sections refine each stage of the framework process presented in Figure 2, including a proof-of-concept example. A trade study of a data recording system is evaluated, with screenshots to help clarify the tools and processes used in each stage.

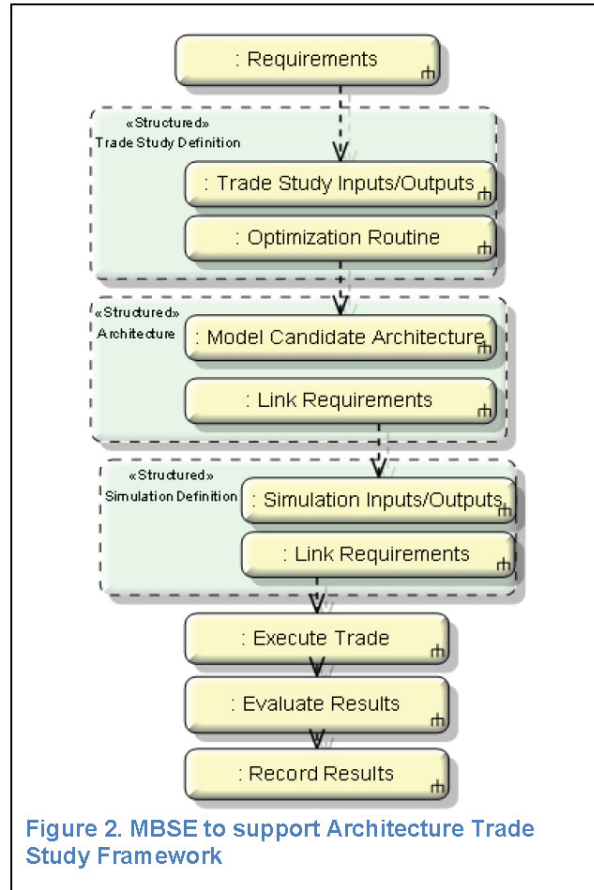


Figure 2. MBSE to support Architecture Trade Study Framework

3.2 System Requirements

Similar to conventional System Engineering Vee process, the initial stage is the identification of stakeholder needs which represent system requirements. System requirements must be gathered and documented, which is often maintained in a requirements repository. In this example we used DOORS which allows for traceability between requirements in terms of associations, derivation and flow-down (see Figure 3). Automatic synchronization allows requirements stored in DOORS to be imported and synchronized to requirements objects in SysML (see Figure 4). SysML offers a requirements diagram to show hierarchy and association of such requirements objects. Additionally these requirements are objects in the model they can be explicitly added to Block Definition Diagrams (bdd) or Constraint Diagrams where they can be linked to identify relationships. This linkage provides

traceability, documentation and extends communication to designers to describe why particular design decisions have been identified.



Figure 3. Requirements captured in DOORS

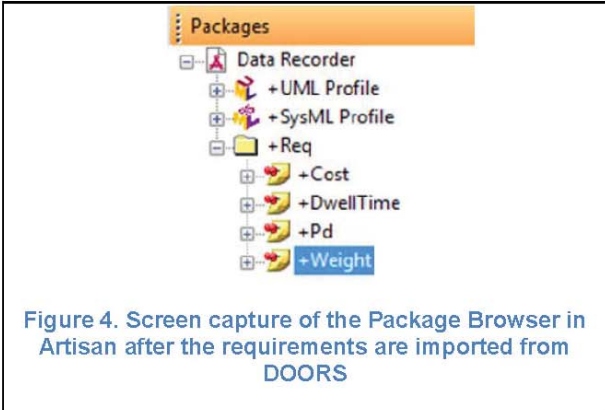


Figure 4. Screen capture of the Package Browser in Artisan after the requirements are imported from DOORS

3.3 Trade Study

Following definition of system requirements, the focus shifts to the motivation and evaluation of the Trade Study. The trade study scope must be identified and documented, encompassing all valid architecture variations and scenario alternatives. From an operational research technique, factors of the trade study can be identified as Objectives, Constraints, Decision Variables or Parameters. Inputs to the study are Decision Variables or Parameters. Requirements can be describes as Objectives (i.e. must maximize dwell time) or Constraints (i.e. must not have a buffer input rate overflow) [16]. This division of requirements allows the optimization routine to select the system that best satisfies the minimum goals of the stakeholders, while giving insight into effective system parameters which can lead to the greatest value. Inline with the Capabilities-Based acquisition trend, this

optimization supports identification of the '80 percent solution' [19].

Optimization Output Parameters	
Objective	Goal
total_cost	min
maxDwellLength	max
Constraint	
bf_InError	FALSE
bf_OverflowErr	FALSE

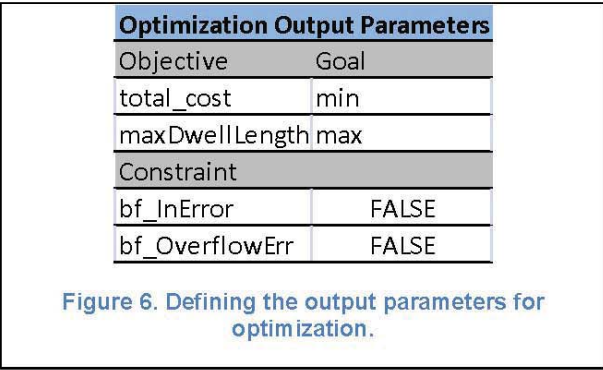
Figure 6. Defining the output parameters for optimization.

Figure 5 shows the inputs to the trade study for this example. Input parameters that can vary in the trade study include the component model type(s), the number of components (i.e. number of buffers can vary from 1-2), and dwell time. Trade study inputs (and supporting parameters associated with the selection of those inputs) are entered into ModelCenter for execution of the trade study.

Trade Space Inputs			
Inputs	Options		
amp	AMP_001	AMP_002	
a2d	A2D_001	A2D_002	A2D_003
buffer	BF_001	BF_002	
nbuffers	1-3		
ssr	SSR_001	SSR_002	
nssr	1-10		

Figure 5. Defining the trade space

Also entered into ModelCenter are the objectives and constraint requirements identified in Figure 6. The DARWIN algorithm optimizer analyzes simulation outputs against set Objectives and Constraints when selecting a final design. In this example we attempt to minimize cost and maximize dwell time (Objectives), so long as a buffer overflow has not occurred or A/D sampler ate to buffer mismatch has not occurred (Constraints).



3.4 Architecture

After identifying what the system needs to do, and how it will be used, potential architectures can be modeled. In an object oriented fashion each block will have critical parameters which can be defined on a per-block basis. Figure 7 is an example of a Block Definition Diagram of the data recorder system. In this viewpoint we want to look at all of the components of the system therefore attribute information is hidden. By examining the system in various levels of detail, each diagram and viewpoint can serve a different purpose. Each distinct viewpoint uses information from a central data repository, thereby establishing consistency amongst objects in the model.

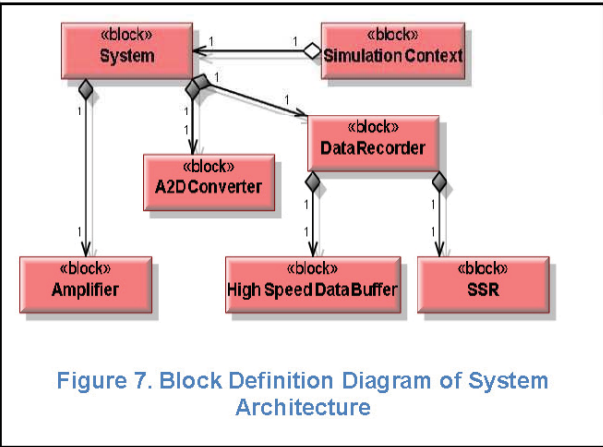
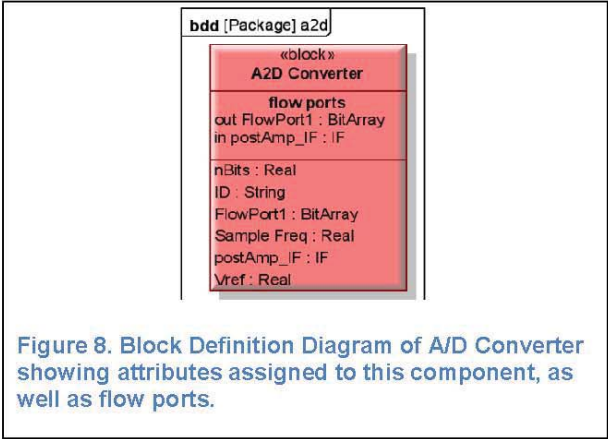


Figure 8 shows a Block Definition Diagram of just the A/D component where all of the attributes are visible. These attributes are abstracted such that each model of A/D given in the Trade Space Inputs can have variable

parameters (ie. A2D_001.nBits = 12 bits, A2D_002.nBits = 18, A2D_003.nBits = 32 Bits). Due to abstraction principles, these instances themselves do not need to individually be modeled, as the associated instance information is documented in the trade study evaluation.



Requirements particularly linked to architecture can be visualized through association, satisfy and refine connections on SysML Block Definition Diagrams (or Parametric Diagrams). Such links allow for traceability when either the block information changes which could affect the ability to satisfy a requirement, or if the requirement changes in which case the selection of a block may need to be re-evaluated.

3.5 Simulation

Once the requirements are understood, the trade space is defined and potential architectures have been identified, simulation models can be designed. Simulation often requires domain – specific tools. The critical foundation for integrating these tools is a defined metadata language, APIs provided by the domain tools or custom interface software. By identifying the simulation inputs and outputs, complex scenarios can be shared amongst a group of engineers with increased communication and interoperability. Identifying requirements associated with each simulation tailors those simulations for specific goals, thereby limiting their scope and

clearly defining the purpose of each simulation. Again, requirements objects can be linked to blocks representing simulations in the model for traceability purposes.

Error! Reference source not found. shows a block diagram of Performance Simulations required in the Trade Study evaluation of the Data Recorder Model (note – this also includes simulations not yet implemented in the framework example). It is obvious from this example that the Performance Simulations can each have an independent focus or work together to fulfill a system requirement.

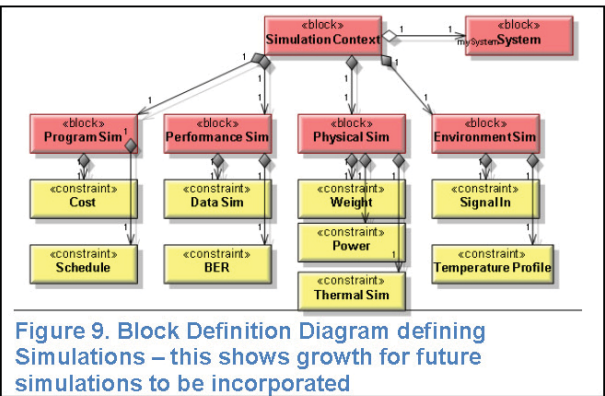


Figure 9. Block Definition Diagram defining Simulations – this shows growth for future simulations to be incorporated

Parametric Diagrams can represent the structure of the Performance Simulation. Inputs and outputs can be identified as ports on Constraints Blocks, and these inputs and outputs can be linked to attribute value fields on the architecture. By first creating a Block Definition Diagram and associating the System Block to the Simulation), then attributes from components in the System Block can be linked to Constraints Block inputs or outputs. In the 'Data Sim' Parametric Diagram figure below requirements are linked, inputs and outputs of the simulation are identified and an example exists of linking a 'A2D Converter' block attribute 'nBits' to an input of the 'Data Sim' performance simulation. Additionally, requirements are linked to the Constraint Block, again using MBSE techniques to enhance communication and traceability.

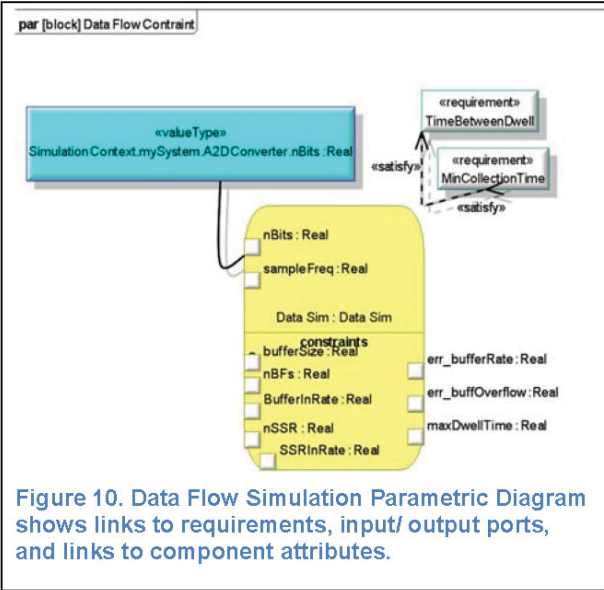


Figure 10. Data Flow Simulation Parametric Diagram shows links to requirements, input/ output ports, and links to component attributes.

3.5.1 Cost Model (excel)

The 'Cost Model' represents a specific example of a Simulation that models the cost of the system based on inputs from the Trade Study and an Inventory Model (Figure 11). Inputs driven by the Trade Study include component selection, compared to an Inventory Model which defines component stock level and cost to compute a final cost (with the added min lot buy complexity). The output computed (

	Inventory			
	In Stock	cost per	min lot buy	order time
AMP_001	1	1000	20	1
AMP_002	2	1500	30	1
A2D_001	0	4000	10	2
A2D_002	3	5000	5	3
A2D_003	2	6000	10	1
BF_001	1	300	2	2
BF_002	0	200	2	3
SSR_001	2	10000	1	4
SSR_002	3	12000	1	5

Figure 11. Inventory model representing components in stock, price, and min lot buy requirements

The output (Figure 12) which will be returned to the overall trade study is the final cost of the system for a given set of selected inputs.

	Inputs		Outputs			
	Part Id	n req	n bought	cost	schedule (mo)	
amp	AMP_002	1	1	\$ 1,500	0	
a2d	A2D_001	1	10	\$ 40,000	2	
buffer	BF_001	2	2	\$ 600	2	
ssr	SSR_001	1	1	\$ 10,000	0	

Figure 12. Example of the output of the cost model given a scenario from the Trade Study inputs

3.5.2 Dwell Time (Matlab)

In this example one of the driving requirements is a given dwell time. A maximum dwell time can be computed based on the output data rate of the A/D, the input rate of the buffer, the size of the buffer memory, the output rate of the buffer and the input rate of the SSDR. Less data from the A/D increases dwell time because not as much data needs to be stored in a given time. Increasing the buffer size increases dwell time because it takes more time to fill the buffer prior to overflow. Increasing the SSDR input rate (or buffer output rate) increases dwell time because the data can flow from the buffer faster, therefore it takes longer for a buffer overflow to occur. In any of these cases, the opposite had a negative impact on dwell time. If the data rate at the output of the A/D exceeds the input rate to the buffer a 'bf_inError' occurs. If the max dwell time computed for a given architecture exceeds the requirement dwell time a 'bf_Overflow Error' occurs. Both errors are output to the Trade Study in ModelCenter (as Constraints) along with the computed 'max dwell time' (as an Objective) which will be used to optimize the dwell time versus cost.

While Matlab and Excel are tools that often communicate, the point of this framework is that by defining the trade study, modelling the simulations ahead of time, properly constraining the simulations, and defining the interdependencies, more complicated simulations can be performed and eventually additional simulations can be incorporated.

3.6 Execute & Evaluate

Finally the trade study is executed on an integrated set of models, and evaluated against the requirements. In this example the simulation models are linked together using ModelCenter. Each execution, Model Center varies the inputs of the trade study, attempting to find the best architecture to meet the needs of the optimizer routine. Again custom software could drive the execution of these simulations in concert, understanding that the driving inputs are generated from the trade study, architecture model and prior simulation output.

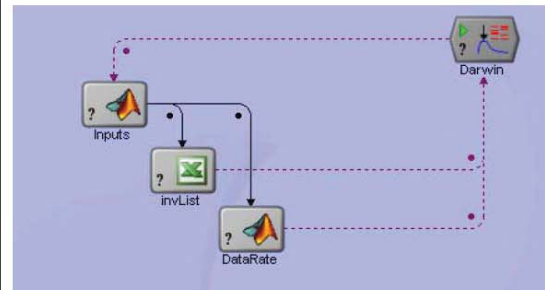
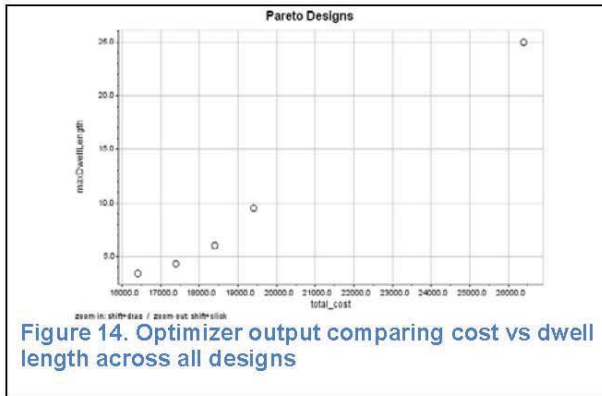


Figure 13. Model Center Project includes Darwin Optimizer, Input definition, Cost Model and Dwell Rate Model

3.7 Analyze Results

The evaluation stage includes assessing the results of the optimization routine. The optimization routine included in ModelCenter Darwin is selected for this example. Decision evaluations such as weighted matrix could also be used to evaluate the trade study results. Figure 14 is an example of the comparison of two Objective requirements (Cost, MaxDwellTime) resulting from input parameter variation. The optimal design minimizes total cost while maximizes dwell time.



After determining the optimum trade study result, (resulting in final architecture definition), the results of the study should be written back into the model for configuration control. ModelCenter displays the selected design (ie. Model.Input. A2D = 2 means that A2D_002 was selected as the converter in the optimal design). This can be supported through a Block Definition Diagram of instances. Again, custom software must be written to integrate the output of the trade study into the SysML models.

Design 1	
Objective(s)	
Model.invList.total_cost	= 16400.0000000
Model.DataRate.maxDwellLength	= 3.4482759
Design Variables	
Model.Inputs.dwell_time	= 0.4000000
Model.Inputs.amp	= 1
Model.Inputs.a2d	= 2
Model.Inputs.bf	= 2
Model.Inputs.n_bf	= 2
Model.Inputs.ssr	= 1
Model.Inputs.n_ssr	= 1
Constraints	
Model.DataRate.bf_InError	= 0.0000000
Model.DataRate.bf_Overflow	= 0.0000000
Watch Variables	

Figure 15. Output of ModelCenter showing design which lead to the optimal requirements analysis

4 DISCUSSION

The framework proposed supports a methodology to integrate MBSE and Performance Simulation to optimize Architecture Trade Studies. Inherent in this framework is the novel approach to modeling simulations which provides definition of

simulation inputs and outputs as well as explicitly linking those simulations to requirements. Benefits of this framework include 1) current configuration of design decision 2) increased communication between system engineers for performance simulation 3) explicit traceability between trade studies and requirements 4) potential for system engineering reuse (including models, hardware, software and analysis tools). Additional case studies will give insight into the scalability of this framework and add fidelity to the integration of SysML directly with Performance Simulations (again tool interoperability). As new tools are developed, and further examples of explicit program application emerge, this framework will provide a guideline for architecture analysis through MBSE and Performance Simulation in the System Engineering user community.

5 REFERENCES

- [1] M. R. Blackburn, "What's model driven engineering (MDE) and how can it impact process, people, tools and productivity," System and Software Consortium (SSCI), Tech. Rep. SSCI-2008002-MC, Sept 2008.
- [2] J. Axelsson. (2002, Model based systems engineering using a continuous-time extension of the unified modeling language (UML). *Systems Engineering* 5(3), pp. 165-179.
- [3] J. Estefan, "Survey of model-based systems engineering (MBSE) methodologies," Jet Propulsion Laboratory, Pasadena, California, U.S.A., 2007.
- [4] INCOSE, "Incase system engineering handbook," 2010.
- [5] A. L. Ramos, "Model-Based Systems Engineering: An Emerging Approach for Modern Systems," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 2010.
- [6] M. Hause, "The Unified Profile for DoDAF/MODAF Enabling Systems of Systems on Many Levels," *IEEE*, 2010.
- [7] Rhodes, D.H., and A.M. Ross, "Anticipatory capacity: Leveraging model-based approaches to design systems for dynamic futures," in *International Conference on Model-Based System Engineering*, 2009, pp. 46.
- [8] R. Rasmussen, M. Ingham and D. Dvora, "Achieving control and interoperability through unified model-based systems and software engineering," Sept 2005.
- [9] J. O. Grady. (2009, Universal architecture description framework. *Systems Engineering* 12(2), pp. 91-116.
- [10] S. Friedenthal, *A Practical Guide to SysML: The Systems Modeling Language*. Morgan Kaufmann, 2008.
- [11] C. J. J. Paredis and T. Johnson. Using OMG'S SYSML to support simulation. Presented at Simulation Conference, 2008. WSC 2008. Winter.
- [12] T. A. Johnson, "Integrating Models and Simulations of Continuous Dynamic System Behavior into SysM," 2008.
- [13] R. Kawahara, H. Nakamura, D. Dotan, A. Kirshin, T. Sakairi, S. Hirose, K. Ono and H. Ishikawa. Verification of embedded system's specification using collaborative simulation

- of SysML and simulink models. Presented at Model-Based Systems Engineering, 2009. MBSE '09. International Conference on.
- [14] D. B. Russell and W. K. McQuay. The joint modeling and simulation system a common modeling architecture for the dod. Presented at Simulation Conference Proceedings, 1993. Winter.
- [15] L. Xing-hua and C. Yun-feng. Design of UAV flight control system virtual prototype using rhapsody and simulink. Presented at Computer Design and Applications (ICCD), 2010 International Conference on.
- [16] A. P. Sage and W. B. Rouse, Eds., *Handbook of Systems Engineering and Management*. Hoboken, NJ: John Wiley & Sons, 2009.
- [17] B. Sauser, J. Ramirez-Marquez, R. Magnaye and W. Tan, "A Systems Approach to Expanding the Technology Readiness Level within Defense Acquisition," *International Journal of Defense Acquisition Management*, vol. 1, 2008.
- [18] United States Government Accountability Office, "DEFENSE ACQUISITIONS: Assessments of Selected Weapon Programs," 2010.
- [19] R. Gates, "Remarks by Secretary of Defense Robert Gates at the Army War College, Carlisle, Pa." *Office of the Assistant Secretary of Defense (Public Affairs)*, .

6 ACKNOWLEDGEMENT

I would like to acknowledge several colleagues who have in some way or another positively contributed this body of knowledge. Rebekah Ceremae assisted in supporting literature reviews. Her ability to find relevant and recent information greatly shaped the definition of the research efforts.. Sean McGervey, who is currently developing a novel the foundation for MBSE efforts at Northrop Grumman guided the research by describing issues facing current MBSE practitioners. Brad Longo assisted through custom tool development, in addition to integration of the performance simulations in Model Center.

2.0 HEALTH & MEDICINE TRACK

Track Chair – Don Combs, Eastern Virginia Medical School



2.1	Interoperability Standards for Medical Simulation Systems.....	137
2.2	Biomedical Simulation Models of Human Auditory Processes.....	143
2.3	Web-Based Predictive Analytics to Improve Patient Flow in the Emergency Department	153

2.1 Interoperability Standards for Medical Simulation Systems

Interoperability Standards for Medical Simulation Systems

Andreas Tolk
Old Dominion University
atolk@odu.edu

Saikou Y. Diallo
Old Dominion University
sdiallo@odu.edu

Jose J. Padilla
Old Dominion University
jpadilla@odu.edu

Abstract. The Modeling and Simulation Community successfully developed and applied interoperability standards like the Distributed Interactive Simulation (DIS) protocol (IEEE 1278) and the High Level Architecture (HLA) (IEEE 1516). These standards were applied for world-wide distributed simulation events for several years. However, this paper shows that some of the assumptions and constraints underlying the philosophy of these current standards are not valid for Medical Simulation Systems. The paper describes the standards, the philosophy, and the limits for medical applications and recommends necessary extensions of the standards to support medical simulation.

1.0 INTRODUCTION

“War is the father of all things!” This statement made by the Greek philosopher Heraclites (520 BC – 460 BC) can be interpreted to be true for the rise of modeling and simulation (M&S). Although M&S has many application domains, from aerospace to Z particle simulation in physics, only the military defense domain applies a broad variety of M&S application in global federations. For example, air defense tanks in Germany provide air protection to mechanized infantry troops in Korea against aircrafts actually flying over Texas. In this case, however, all participants are brought together in a virtual battle sphere, not in a real battlefield. In this virtual battle sphere they interact with each other like they would in real systems in a real battlefield.

Two international simulation interoperability standards were developed by the M&S community to support such world-wide simulation endeavors, the IEEE 1278 standard family defining the *Distributed Interactive Simulation* (DIS) protocol, and the IEEE 1516 standard family defining the M&S *High Level Architecture* (HLA). There are other solutions that seek to improve

interoperability, like the *Training and Testing Enabling Architecture* (TENA).

These developments have been applied with some success by the military. It is therefore worthwhile to look at their applicability in support of medical simulation systems. The objective of this paper is to show that this is not possible without ensuring that very strict and rigorous constraints are fulfilled, namely, to ensure that there is one common view of the problem. This is easier said than done as a common view requires that conceptualizations are shared by all models and simulations.

To prove this point, section 2 presents a short overview of the main characteristics of current simulation interoperability standards is provided. Section 3 shows the limits of their applicability in medical simulation. We show that an “additional dimension” is required that is not yet captured in current standardization efforts. Finally, the paper concludes that although many good ideas can be readily transferred from the military domain to the medical, additional research efforts are needed to ensure that simulation standards support valid federated compositions from valid simulations.

2.0 SIMULATION INTEROPERABILITY STANDARDS

This section will summarize the main ideas underlying the two IEEE simulation interoperability standards: DIS and HLA. It also describes the main ideas of TENA and summarizes some of the current LVC approaches.

Initiated in 1983 by the Defense Advanced Research Projects Agency (DARPA), the SIMulator NETworking (SIMNET) project was the first effort exploiting the new developments in computer communications and network technology for simulation. The project emphasized tactical team performance by bringing together armor, mechanized infantry, helicopters, artillery, communications, and logistics. All these simulation systems shared a common battle space. Entities did move, shoot, look, and also communicated. The interactions of these entities focused on physical aspects. Although the systems differed in resolution and details, they shared both a common understanding of the problems to be solved and the worldview underlying the battle space [1]. SIMNET is the common ancestor of DIS, HLA, TENA, and LVC efforts.

2.1 IEEE1278 – DIS

SIMNET was generally perceived to be a success. Tactical teams could now train together, fighting "shoulder to shoulder" in a virtual battlefield. Industry took this idea and turn it into a standard to support distributed interactive simulation. This standard is based on the idea that each system represents one or several individual entities that act based on common rules – the Newtonian laws of physics – in a common battle space. Interactions between these entities could be identified, categorized and standardized.

As a result, individual simulators representing weapon systems on this common virtual battlefield have a well defined set of actions and interactions: tanks could move, observe, shot at each other, exchange radio communication, etc.

Individual activities led to status changes that were communicated via status reports. Interactions were communicated via messages. If two tanks engaged in a duel, the order of activities and the data to be exchanged between these entities were well defined.

Therefore, the set of information exchange specifications could be well defined. This resulted in standards for these messages, which led to the IEEE1278 Distributed Interactive Simulation (DIS) standard [2]. The Protocol Data Units (PDUs) captured (syntactically and semantically) all possible actions and interactions based on the idea that individual simulators represent individual weapon platforms.

The continuous success of DIS was based on a simple concept: individual entities following the laws of physics while exchanging messages. Messages, in this case, represent interactions.

2.2 IEEE 1516 – HLA

DIS, however, has a limitation: it is hard to apply to aggregated simulations or to simulation paradigms other than event driven simulation. The HLA interoperability standard [3] was developed to maximize the flexibility for all kind of M&S application domains and to support different M&S paradigms.

In HLA, ten rules guide information exchange between simulation systems (called federates) and a common infrastructure and network (called Runtime Infrastructure -RTI-). The resulting system is called a federation.

The information exchange requirements within a federation are captured in the Federation Object Model (FOM). The FOM defines all persistent objects and their attributes and transient objects and their parameters that can be exchanged between participating simulations. While persistent objects have to be created and updated in the participating simulation systems during

runtime, transient objects are created in case of need and only used once.

Because of synchronization challenges, six service groups ensures that need information – but only that information that is needed – is delivered at the right time to the right simulation system. The six service groups are federation management, declaration management, object management, data distribution management, time management, and ownership management.

The purpose of *Federation Management* is to determine the federation. Federates join and leave the federation using the functions defined in this group. *Declaration Management* identifies which federate can publish and/or subscribe to which information exchange elements. This defines the type of information that can be shared. *Object Management* manages the instances of shareable objects that actually are shared in the federation. Sending and receiving and updating belong into this group. *Data Distribution Management* ensures the efficiency of information exchange. By adding additional filters this group ensures that only data of interest are broadcasted. *Time Management* synchronizes the federates and *Ownership Management* enables the transfer of responsibility for instances or attributes between federates.

HLA significantly increased the flexibility of simulation federation definitions. Instead of being limited to predefined information exchange groups, the developer can specify the objects and interactions and can even support different time model philosophies. It neither assumes the level of resolution – HLA supports component level simulation, platform level simulation, and all levels of aggregation – nor does HLA assume the partition of the battle space into tactical unit or the phasing of a supported operation.

However, the underlying philosophy remained the same: all simulators are

based on the same worldview of focusing on technical-physical aspects.

2.3 TENA and LVC

The Test- and Training Enabling Architecture (TENA) [4] is another successful solution predominately used on the test ranges for Test- and Evaluation (T&E). TENA sets up a framework enabling the reuse of components and the integration of independently developed systems into a federation. However, the philosophy and design differs. TENA supports the integration of HLA and DIS based systems, but is neither HLA- nor DIS-based. TENA is not standardized by an international standardization organization, but is a de facto standard in the domain of T&E.

Following the TENA philosophy, interoperability requires a common architecture (which is TENA), the ability to meaningfully communicate, which requires a common language (TENA Object Model), and a common communication mechanism (TENA Middleware and Logical Range Data Archive). It also requires a common context (in the form of a common understanding of the environment), a common understanding of time (provided by TENA Middleware), and a common technical process (provided by TENA processes). Similarly, practical reuse and composability require the above, plus well-defined interfaces and functionality (provided in the form of tools and repositories), and a place to store reusable components (repository). TENA provides all these capabilities.

The central TENA object model is a real object model with semantically well-defined objects. As these are real object-oriented object, each object is not only defined by its data, but also by its potential behavior (in the form of methods that can be executed by the objects). This approach combines DIS and HLA ideas: the middleware provides functionality comparable to the RTI of HLA, and the TENA objects are as well defined as the DIS PDU. Ambiguities are

therefore minimized – but for the price to have to use objects of TENA. If new effects or objects are needed, an extension of TENA is required before it can be handled.

Generalizing the ideas coped with so far, the recent developments for Live-Virtual-Constructive (LVC) federations emerged. Live simulation uses real people and real systems in a simulated operation. Virtual simulation uses real people and simulated systems in a simulated operation. Constructive simulation use simulated people and simulated systems in a simulated operation. As such, LVC tries to establish a framework that helps to federate live systems, such as tanks and airplane, with simulators, such as tank and flight simulators, with simulation systems. The resulting architecture shall not have the limits of current domain or paradigm specific solutions. A recent study [5] did not come up with technical solutions, but at least with a common vision and recommended first steps to prepare for such a solution.

In summary, the interoperability standards developed predominantly by the military and related solutions were successfully applied. They enable nation- and world-wide simulation federations and increase the decision quality significantly. It is therefore a worthwhile effort to check their applicability in other domains as well.

3.0 LIMITS FOR MEDICAL SIMULATION

How can medical simulation benefit from the success stories of military simulation? Is it possible to simply adapt IEEE1278 or IEEE 1516 to the needs of this new user community? In particular IEEE1516 – the High Level Architecture – was successfully applied in other domains, such as transportation or even supporting NASA simulation efforts.

In the following section, we will look at the underlying philosophy of the approaches discussed in section 2 and their challenges.

3.1 Philosophy of Current Simulation Interoperability

As discussed in more detail in [6] and as mentioned on the sideline in section 2, the underlying philosophy of current simulation interoperability is dominated by reductionist approaches based on positivism. The technical-physical world of shooting, moving, looking, and communicating of weapon systems supports such a viewpoint. For instance, a simulated weapon system has a one-to-one correspondence to a weapon systems in a real battle space and jointly correspond to one common view of reality. There exists one world and one truth, and it is possible to find this truth by observation and experimentation.

However, since the early 20th century, this view shifted with relativity, quantum physics, and other scientific insights. Science started to shake the believe of one world and one reality that can be observed and understood by reductionism. Furthermore, social sciences contributed as well to move away from the one-worldview assumption. Interpretivism holds the belief that truth is a construct of the observer. Reality is relative and cannot be separated from the individual who observes it. The majority of social and human sciences subscribe to interpretivism.

Modeling actually can be a perfect example for this philosophy. In a simulation, reality is not necessarily what is real, but what is captured in a model. Something is true when it is true in the model. This means that the implemented system can derive the truth value through computation. Two simulation systems are interoperable when they derive consistent truth values; two models are composable when they represent the same worldview.

Interoperability of simulations requires composability of models, and current simulation standards do not take the model level into consideration, they exclusively focus on the implementation level [7].

3.2 Multi-Scope, Multi-Resolution, and Multi-Structure

Another aspect that is not sufficiently addressed by standards, despite of a large body of knowledge compiled by academia [8], is the challenge of multiple scopes of a model, multiple resolutions, and multiple structures.

If two model use different referent systems of their perception, we deal with a challenge of differences in scope. If two models use the same referent system, but they use different attributes with different granularity, they differ in resolution. If two model use the same attribute but compose these attributes into different entities, the use different structures. All three challenges can be observed at the same time, making dealing with these challenges hard.

3.3 Problem Situations

When applying M&S, we assume the existence of a well-defined problem for which there is a solution that needs to be found. However, in particular in soft science – and often observed in the medical domain, this is not necessarily the case. One expert sees a problem, another expert perceives that everything is normal. This does not necessarily mean that one is right and the other wrong, it may simply be that they follow different theories.

To deal with this kind of situations, the term problem situation has been coined [9]: Problem situations are problems for which there is no agreement on the nature of the problem or whether there is even a problem.

4.0 CONCLUSION

The current simulation standards IEEE1278 and IEEE1516 have been successfully applied for military systems when applied in support of technical-physical models. These models all support one worldview (Newtonian physics) and support common sponsors' needs that establish well understood problems.

If a domain does not subscribe to a reductive positivistic philosophy, the application can be problematic, as the standards do not address challenges on the modeling level. As stated in [10]: *"While modeling is the process of abstracting, theorizing, and capturing the resulting concepts and relations in a conceptual model, simulation is the process of specifying, implementing, and executing this model. Modeling resides on the abstraction level, whereas simulation resides on the implementation level. Simulation systems are model-based implementations. Whether or not it is possible to merge two simulation systems in collaborative support of user requirements not only depends on the integratability of the underlying networks and the interoperability of the simulation implementations, but also on the composability of the underlying models."*

In order to support composability of medical models and interoperability of medical simulation, the challenges have to be addressed:

- considering problem situations instead of assuming well-defined problems,
- allowing and identifying different viewpoints (interpretivism) instead of one common worldview (positivism), and
- better support of models with multiple scope, resolution, and structures.

The current simulation interoperability solutions work well for the specified special case, but are unlikely to be equally successful on the broad scale over the complete spectrum.

As stated in [6]: *"It would be naïve to apply standards that were developed for physical-technical models based on a common theory representing the positivistic worldview to integrate socio-psychological models derived from competing theories representing interpretivism and expect valid*

results.” The same is true for the medical domain.

Instead of shoehorning simulation into a federation, the alternative may be to work towards a domain of new standard efforts: the efficient and effective support of *exploratory analysis under uncertainty and disagreement*, and supporting development of strategies that are flexible, adaptive, and robust. Recommendations given in [8] and [11] towards the use of multi-simulations should be researched for medical simulation systems as an alternative to the current solutions that are too limiting and only valid in a small spectrum of the interest domain of medical simulation.

5.0 REFERENCES

- [1] Margaret Loper and Charles D. Turnitsa: History of Combat Modeling and Distributed Simulation. In *Engineering Principles of Combat Modeling and Distributed Simulation*, edited by Andreas Tolk, Wiley, Hoboken, NY (in press).
- [2] Institute of Electrical and Electronics Engineers. *IEEE 1278 Standard for Distributed Interactive Simulation*, IEEE publication, Washington, DC.
- [3] Institute of Electrical and Electronics Engineers. *IEEE 1516 Standard for Modeling and Simulation High Level Architecture*, IEEE publication, Washington, DC.
- [4] Edward T. Powell, Jason Lucas, Kurt Lessmann, and George J. Rumford: The Test and Training Enabling Architecture 2002 Overview and Meta-Model, in *Proceedings of the Summer European Simulation Interoperability Workshop*, 03E-SIW-041, 2003.
- [5] Margaret L. Loper, and Danny Cutts: *Live Virtual Constructive Architecture Roadmap (LVCAR) Comparative Analysis of Standards Management*. Final Report, US M&S CO Project No. 06OC-TR-001. 2008.
- [6] Andreas Tolk, Saikou Y. Diallo, Jose J. Padilla, and Charles D. Turnitsa: How is M&S Interoperability different from other Interoperability Domains? In *Proceedings of the Spring Simulation Interoperability Workshop*, 11S-SIW-008, 2011.
- [7] Andreas Tolk: M&S Body of Knowledge: Progress Report and Look Ahead, *SCS M&S Magazine*, Vol. 1, No. 4, October 2010.
- [8] Andreas Tolk, Saikou Y. Diallo, Jose J. Padilla, and Heber Herencia-Zapana: Model Theoretic Implications for Agent Languages in Support of Interoperability and Composability. In *Proceedings of the Winter Simulation Conference 2011*, Phoenix, AZ.
- [9] Jac A. M. Vennix: *Group Model Building. Facilitating Team Learning Using System Dynamics*. Wiley: Chichester, 1996
- [10] Patrick T. Hester, and Andreas Tolk: Applying Methods of the M&S Spectrum for Complex Systems Engineering. In *Proceedings of the Emerging Applications of M&S in Industry and Academia (EAIA) Spring Simulation Multiconference*, 2010, Orlando, FL, pp. 17-24
- [11] Levent Yilmaz, Tuncer Ören, Alvin Lim, and Simon Bowen: Requirements and Design Principles for Multisimulation with Multiresolution, Multistage Multimodels. In *Proceedings of the Winter Simulation Conference, 2007, Washington, DC*, pp. 823-832

2.2 Biomedical Simulation Models of Human Auditory Processes

Biomedical Simulation Models of Human Auditory Processes

Mehmet M.A. Bıçak, Ph.D
Adaptive Technologies Inc.
mehmet@adaptivetechinc.com

Abstract. Detailed acoustic engineering models that explore noise propagation mechanisms associated with noise attenuation and transmission paths created when using hearing protectors such as earplugs and headsets in high noise environments. Biomedical finite element (FE) models are developed based on volume Computed Tomography scan data which provides explicit external ear, ear canal, middle ear ossicular bones and cochlea geometry. Results from these studies have enabled a greater understanding of hearing protector to flesh dynamics as well as prioritizing noise propagation mechanisms. Prioritization of noise mechanisms can form an essential framework for exploration of new design principles and methods in both earplug and earcup applications. These models are currently being used in development of a novel hearing protection evaluation system that can provide experimentally correlated psychoacoustic noise attenuation. Moreover, these FE models can be used to simulate the effects of blast related impulse noise on human auditory mechanisms and brain tissue.

1 INTRODUCTION

Modeling the performance of hearing protection devices (HPDs), using numerical methods such as Finite Element (FE) Analysis techniques, is needed for further improving the noise attenuation performance of hearing protectors for personnel who work in high noise environments. A typical example for such environment is an aircraft carrier deck, where noise levels can range from 130 dB sound pressure level (SPL) to as high as 150 dB SPL due to exhaust gases emitted from aircraft engines. Wearing an effective HPD is critical in this acoustically dangerous environment to prevent permanent hearing loss. Conventional HPDs are designed to reduce the risk of hearing damage from acoustic energy transmitted to the eardrum through air conduction (AC) through the ear canal. However, the maximum level of hearing protection obtained by using a conventional HPD is limited due to the presence of flanking paths of energy transmission to the cochlea. This flanking energy transmission is referred to as “bone conduction” (BC). It is accepted that the majority of the transmitted energy via BC to the middle and inner ear components is through the cartilage, skin, soft tissue and fluids of the human head. The limit of hearing protection performance set by the BC pathways is commonly termed the “BC limit” in hearing protection literature [1]. There has been increased development of computer simulations in the area of auditory

mechanics [2,3,4,5], which has been facilitated by the growing sophistication of hardware and software technology. These simulations allow researchers to explore complicated biomechanical systems, often times involving numerous components with intricate geometries, complicated non-linear material properties and challenging boundary conditions. FE simulation methods offer significant advantages over traditional experimental methods, partly due to the inability of experimental methods to effectively measure internal human head and auditory system response to acoustic excitation. In contrast, all relevant structural and acoustic information, such as mode shapes, displacements, pressures and energy flow (intensity), are accessible at any arbitrary location within the FE simulation. FE simulation also offers the ability to visualize and study the three-dimensional dynamic response of the system, allowing for the ability to gain significant insight into the dynamic system behavior. However, the usefulness of the FE simulation is largely dependent on the fidelity of the model representative of the physical system. Therefore, accurate and relevant experimental information is needed to support the development of FE simulations.

This study, the use of experimental data to develop a FE simulation of a human subject for the investigation of AC and BC pathways, as well as HPD to flesh dynamics, is presented in this paper.

2 AUDITORY PATHWAYS

Understanding AC and BC phenomena is critical in developing improved hearing protection devices (HPDs) for extremely high noise environments. However, its underlying mechanics, especially with respect to acoustically induced BC, are not sufficiently understood. This is in part due to the relative complexity in developing analytical and/or numerical models that accurately capture the interactions between the macro dynamics of the skull and the micro dynamics of the middle ear and cochlea. Figure 1 shows the components of the AC and BC pathways which include the pathways presented by Stenfelt [6] as well as the pathways due to the presence of an earplug.

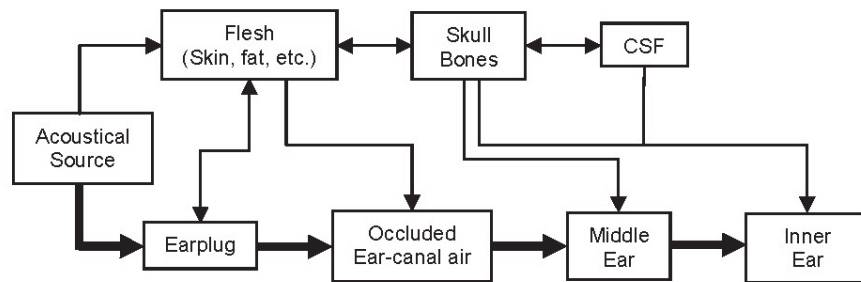


Figure 1. A model of acoustical pathways considering the occlusion due to the earplug. Occluded AC pathway is depicted by using thick lines.

When the external ear canal is effectively occluded by a high attenuation HPD, sound may not reach the cochlea through the direct acoustic excitation of the eardrum. In this case, sound energy may be transmitted to the middle-ear and the inner ear region via:

- Acoustically-induced skull vibration
- Transmission through nasal/oral cavities and the Eustachian tube to the middle ear
- Residual vibration of the earplug and canal bone-skin layer

Traditionally, most researchers in the BC field have focused on the skull vibration path (both rigid-body and flexural vibration) [7,8,9] because this is the most intuitive path of energy transmission to the auditory system, and because development of a BC exciter for medical purposes is often a

research focus. Studies show that the average human skull *in vitro* exhibits the first resonant frequency at around 1 kHz [10]. Therefore, when subjected to external acoustic or structural excitation, a skull will exhibit dominant rigid-body vibrations (i.e. translation and rotation) at frequencies below about 1 kHz and flexural vibrations at frequencies above 1 kHz. In the current context of hearing protection, which involves an acoustic excitation rather than direct structural excitation, it is also possible that the other two paths described above may contribute to BC at certain limited frequency ranges.

Following the acoustic energy transmission through the skull to the auditory system, the main mechanisms involved in the excitation of the basilar membrane (BM) vibration in the cochlea are often categorized as follows:

- BC response of the middle ear
- BC response of the inner ear

Although separating the BC mechanisms into these two categories for simplification has been customarily done in literature, it should be noted that the mechanics of the middle ear and the inner ear are inter-related and one cannot completely treat them as independent.

3 DEVELOPMENT OF FE MODEL

To achieve the goal of developing a FE model which is capable of simulating AC and BC sound transmission, it is important to investigate human anatomical features

separately. It is impractical to model all the physiological features present in the actual physical system. Therefore, a certain degree of simplification has to be carried out, selecting only the essential features of the actual physical system for modeling based on a *priori* knowledge.

The high-fidelity FE models of the auditory system, including the relevant HPD, should be created using accurate geometrical information obtained from 3D imaging data. However, only skull, outer ear, ear-canal and the locations of middle ear bones can be extracted using a volume computed tomography (CT) scanning due to current limitations of spatial resolution from the lower exposure levels. In order to obtain highly detailed geometrical information, microCT scanning may be employed. However, this method is only applicable to cadavers due to intense dose levels at the human head. Due to this issue, FE model middle ear and cochlea geometries of a specific subject are usually based on cadaver scanned geometry. The developed FE models of the skull and outer-ear from CT scans as well as the isolated middle-ear and cochlea FE model based on microCT scans should be integrated to form a complete geometrical part of the whole auditory system model. An example of this process is presented for the middle-ear and cochlea in Figure 2 and 3.

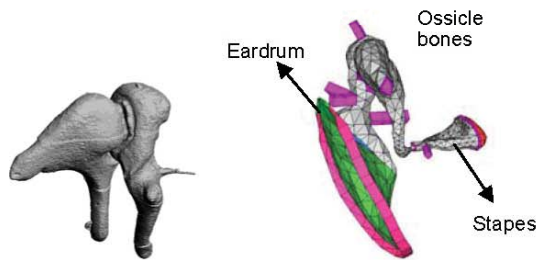


Figure 2. 3D solid model of malleus and incus derived from microCT imaging data (left), middle-ear FE model (right).

The integration of these two FE models with a developed skull model will result in having the complete FE model which is shown in Figure 4.

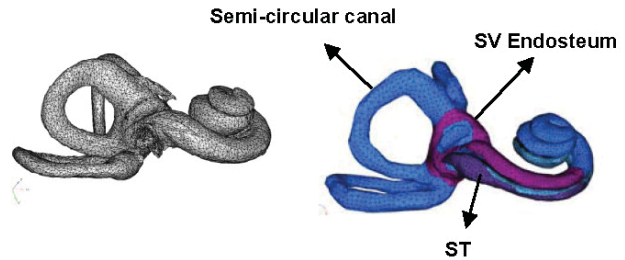


Figure 3. 3D solid model of cochlea derived from microCT imaging data (left), FE cochlear model with SV¹ endosteum and ST² (right). (Not to scale)

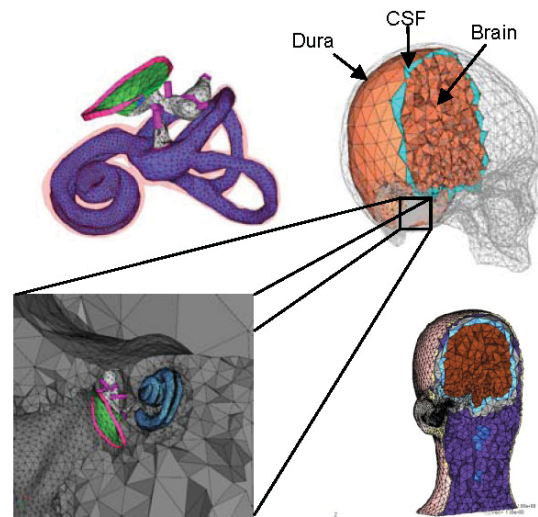


Figure 4. Integrated FE models, cochlea-middle ear (topleft), outer-ear skull including cochlea-middle ear FE (topright) and whole head FE model (bottomright).

4 NUMERICAL RESULTS AND VALIDATION

Material properties of various components should be selected by consulting the data provided in existing references on auditory system FE modeling [2,3,11,12]. The past literature reveals that large uncertainty still exists regarding the values of material properties for auditory system components. In past auditory system FE modeling efforts, material property values were typically determined by iteratively adjusting the literature values until good correlation was obtained between simulated and experimental dynamic response data [13]. This approach is adapted for the current FE

¹ Scala Vestibuli

² Scala Tympani

model in this study; the different material property values presented in literature are iteratively updated until agreement was obtained between published experimental data of dynamic component response and FE model results.

4.1 Simplification of circum/intra-aural flesh

One important simplification of the FE model is to use a single isotropic material for modeling the internal physiological structure of the human circum/intra-aural flesh (which includes the ear cartilage, fat, muscle and skin) rather than modeling the components individually. The validity of this assumption is investigated by comparing two FE models. One model is a realistic multi-component flesh model that contains the individually modeled internal structural components, while the second model is a homogenized flesh model, where the flesh is assumed to consist entirely of one type of material.

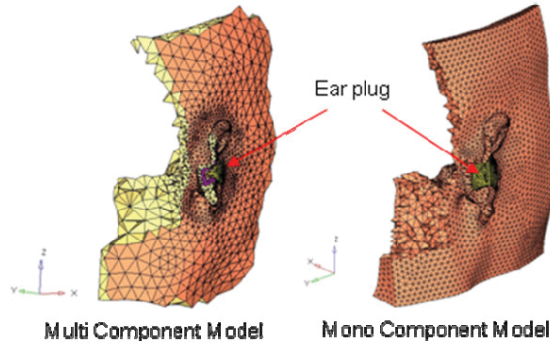


Figure 5. Multi-component vs. homogenized circum/intra-aural FE models.

Figure 5 shows images of these two models. Figure 6 shows the simulated insertion loss results for the multi-component model versus the homogenized flesh model results, along with experimental data. Four separate cases for the homogenized model are created with differing values for the flesh material elastic modulus; the values are increased by a factor of 1, 1.5, 2 and 3. It is observed from Figure 6 that the insertion loss for the case with a homogenized flesh, with the modulus increased by a factor of 2 from the baseline, shows good agreement with the baseline

multi-component model. This suggests that the flesh model can be simplified into a single material structure with an effective elastic modulus, to ultimately reduce model complexity.

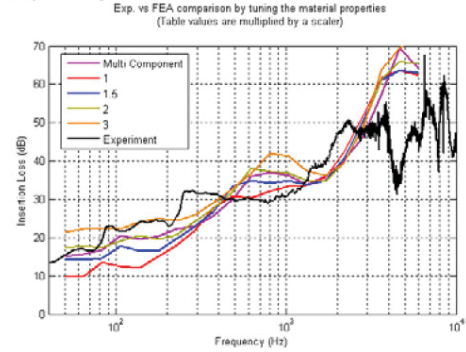


Figure 6. FE-simulated earplug attenuations (insertion loss) for the multi-component and homogenized circum/intra-aural flesh model. The Young's modulus of the flesh material for the homogenized flesh model was increased by a factor of 1, 1.5, 2 and 3.

4.2 Skull model

Mechanical response of the skull FE model with cranial content is compared against published experimental data [10,14]. The current skull model results with cranial content are in agreement with experimental data of point mechanical impedance at a location around mastoid, as shown in Figure 7. Figure 8 shows acceleration at contra- and ipsi-lateral cochlea locations in response to a force input at the mastoid location.

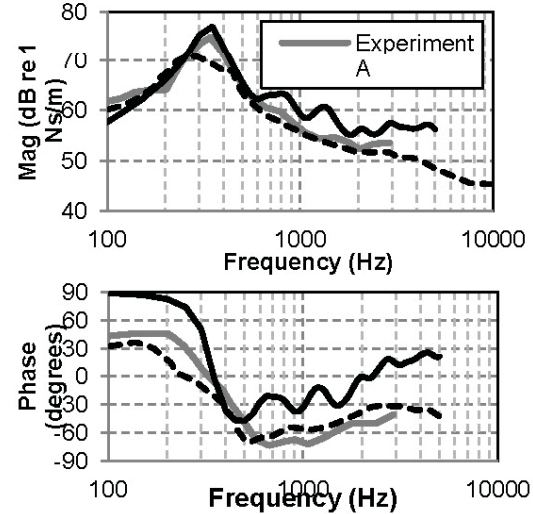


Figure 7. FE and experimental point mechanical impedance response for a skull with the cranial content.

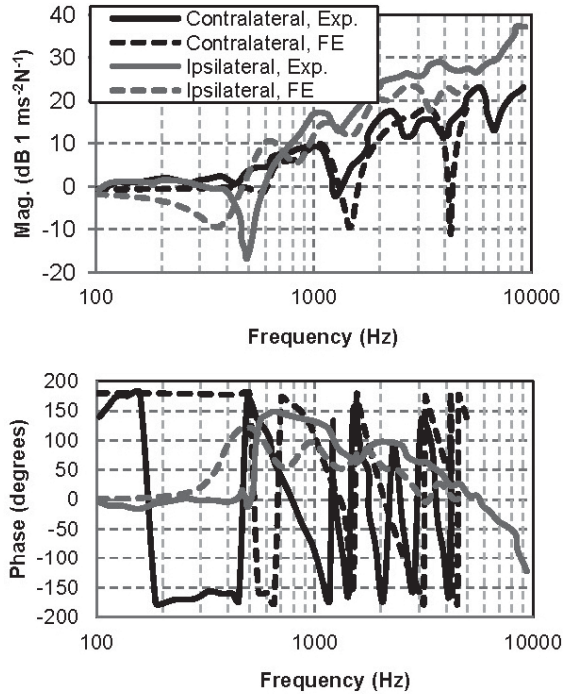


Figure 8. FE and experimental ipsilateral and contralateral acceleration response for a skull with the cranial content.

4.3 Middle-ear and Cochlea models

The FE model material properties and mesh characteristics of the middle-ear were discussed previously in Homma et al. [15]. The BM, which is determined to be the most significant component in the cochlea, is developed with orthotropic material properties. Detailed mechanical property and geometry of the BM were also described in the final report for an Air Force STTR program, topic number AF06-T035 [13]. To combine the middle-ear and the cochlear models, the stapes is attached to the cochlea via the stapes annular ligament, while the nodes along the perimeter of the stapes annular ligament are fixed to the cochlear bony shell. The current middle ear FE model does not include the air volumes associated with the middle ear cavity and the ear canal, primarily because the loading effects due to the air volumes are not considered to be significant under normal conditions. However, when air volumes are made significantly small by deep-occlusion of the ear canal or by a significant reduction in middle ear cavity volume, it is possible

that the reduced air volume could provide additional stiffness to the middle ear system, therefore altering the system response characteristics. If this proves relevant, then it is likely that air volumes must be included in FE models. However, the reference experimental data of the middle ear, in both AC and BC conditions, are obtained without significant air loading, therefore this assumption is likely to be valid.

Model parameters are adjusted so that the FE model exhibits the correct impedance of the stapes/cochlea system with the fluid-filled cochlea, as well as with the cochlear fluid removed. Figure 9 shows that the impedance contribution from the cochlea itself is mostly resistive, which is evident from the increased damping effect around the system resonance at 3-4 kHz.

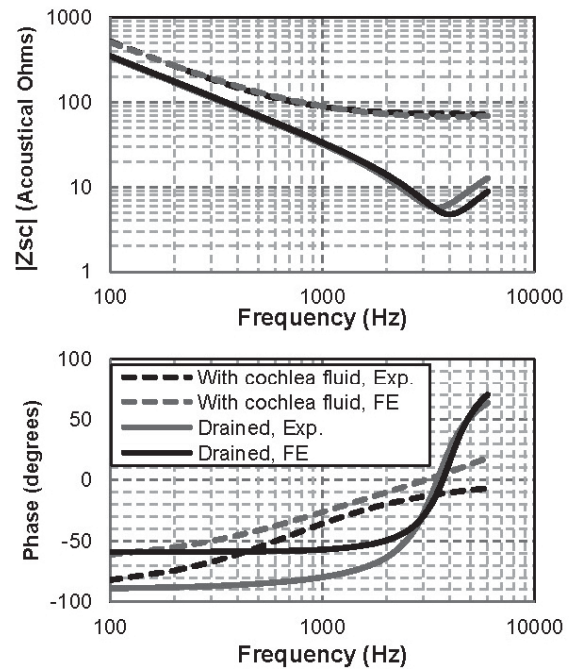


Figure 9. Input acoustic impedance (pressure / velocity) of cochlea/stapes system, Z_{sc} : matching FE model response against target response data [16].

The middle-ear pressure gain function, which is defined as the ratio of oval window pressure (P_{OW}), to ear-canal pressure (P_{EC}), is another useful metric for FE model

validation, and is shown in Figure 10 and Figure 11. In this validation study, P_{EC} is represented by the unit pressure applied on the tympanic membrane (TM).

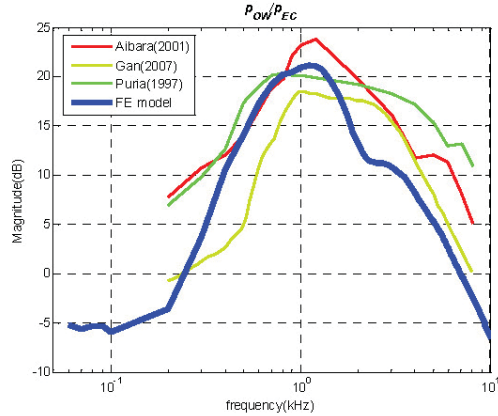


Figure 10. The ear-canal pressure magnitude, p_{EC} , to oval-window pressure, p_{OW} , middle-ear gain.

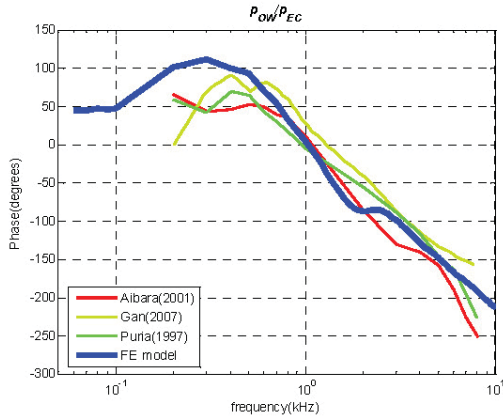


Figure 11. The ear-canal pressure phase, p_{EC} , to oval-window pressure, p_{OW} , middle-ear gain.

P_{OW}/P_{EC} is consistent with experimental data within a 5 dB range from 300 Hz to 10 kHz. Figure 12 shows simulated cochlear input impedance (the ratio of the scala pressure near the oval window (OW) to the volume velocity at the OW, P_{OW}/U_{OW}) compared with published data. Figure 12 indicates that the FE result is qualitatively consistent with published data [10,16,17,18]. However, it is observed that the simulated result is less damped or over-massed than the experimental data.

Figure 13 shows the FE-simulated Best-Frequency (BF) map along with those obtained experimentally by Greenwood [19]. The BF map indicates the peak BM

vibration location among 200 measurement points along the BM as a function of frequency.

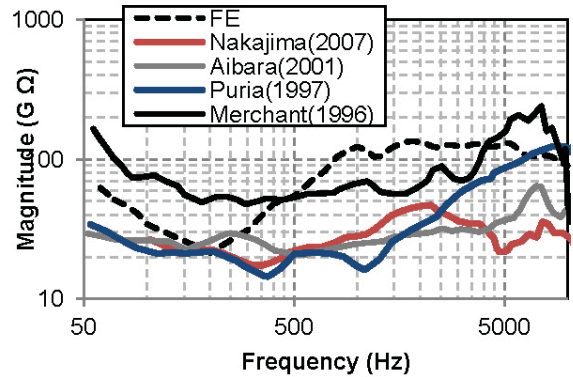


Figure 12. Cochlear input impedance, p_{OW}/U_{OW} . Magnitude in $G\Omega$.

As shown in Figure 13, the BF map from the FE model is in agreement with the experimental data. This agreement was a result of an iterative tuning process of elastic modulus values of the BM with AC excitation.

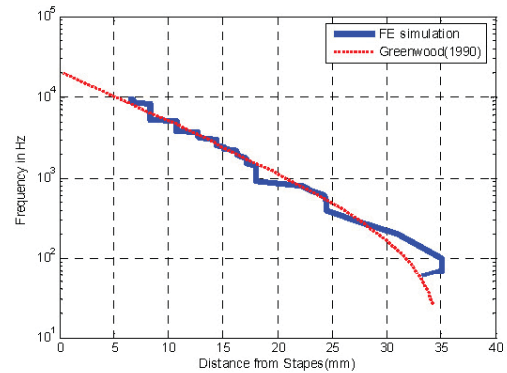


Figure 13. Simulated BF map for AC excitation along with experimental data by Greenwood [19].

Figure 14 shows the simulated and measured AC response of the BM velocity, v_{BM} , observed at a specific BM location (12 mm from the base) normalized by the stapes footplate (i.e. OW) velocity, v_{OW} . Since both BM velocity and stapes footplate velocity are vectors whose components are complex values, the calculated value with two vectors, v_{BM}/v_{OW} , cannot give any physical meaning of normalized velocities in the respect of magnitude and phase.

Therefore, the normal vectors of BM and stapes measurement points, n_{BM} and n_{ST} are obtained. It should be noted that the positive direction of each normal vector in this study is the direction from the ST to SV for the BM, and the direction from the outer ear to the inner ear for the stapes. Next, the inner product between the velocity vector and the normal vector of the measurement points is performed to obtain complex scalar values, which were the normal directional components of the velocity vector for the BM and stapes measurement points. The normalized BM velocity is calculated by the normal component of the BM velocity over the normal component of the stapes velocity. The FE result in Figure 14 shows a reasonable agreement with the measured results from the literature in terms of its overall response characteristics.

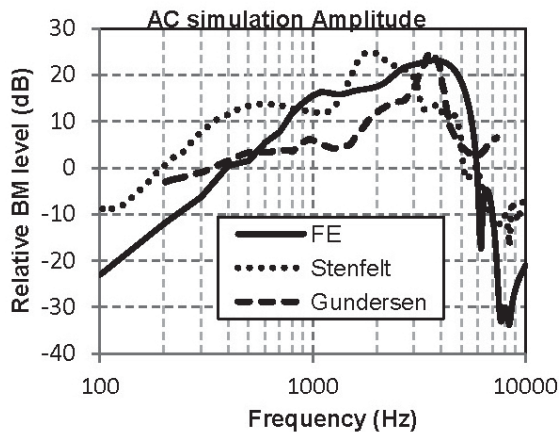


Figure 14. Auditory system model validation: (a) middle-ear pressure gain (scala vestibuli pressure over TM pressure) and (b) BM velocity in response to TM pressure normalized by stapes velocity. BM velocity was observed at 12 mm from the base for the FE result. The cadaver ear experimental data [20,21] were also obtained at a similar location on the BM.

5 VIRTUAL EXPERIMENTS

Having a FE model correlated against literature data gives a unique opportunity to further investigate the challenging questions in AC, BC and resulting pathways. The AC and BC limit difference as well as

transmission paths due to the earplug are investigated.

5.1 FE Model Simulation of AC/BC Difference ("BC limit")

The difference in magnitude of AC perceived sound to BC perceived sound is referred to as the "bone conduction limit", which suggests the limit of hearing protection achievable by occluding the ear. This effect is simulated by computing the difference between the BM vibration levels, as indicated in terms of BM structural deformation for AC and BC cases, with results presented in Figure 15. The AC case is simulated by applying acoustic pressure on all external surfaces of the skull model, including the TM. This result is then compared with the BC case, where the pressure was again applied over the external skull surfaces, but not on the TM. The simulated AC/BC level difference in Figure 16 is presented to be similar to published data [1,22,23], despite the fact that the simulation lacked the flesh layer covering the skull.

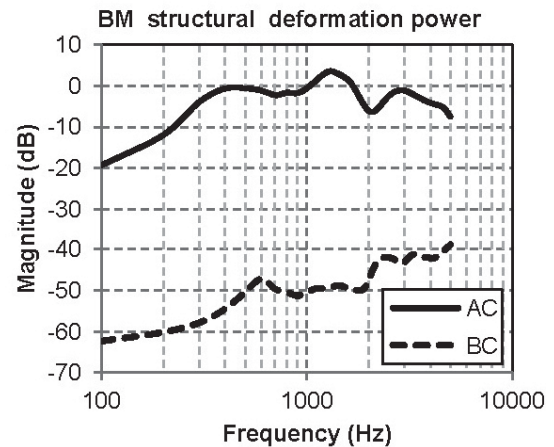


Figure 15. Simulated AC/BC sound transmission difference (i.e. BC limit) obtained with the current model: deformation power of the vibrating BM for AC and BC excitations.

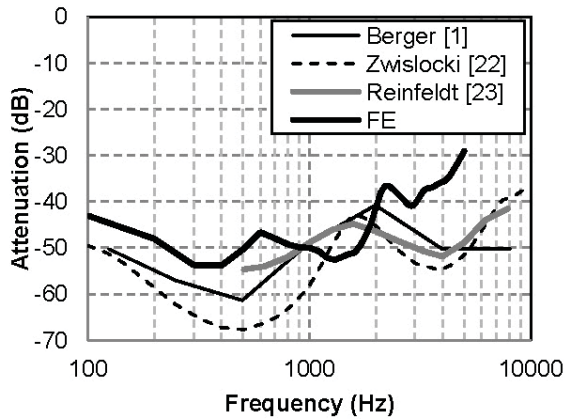


Figure 16. FE-simulated BC limit (dB difference in BM vibratory deformation power between AC and BC) compared with published data [1,22,23].

The motion of the cochlea structure as the skull undergoes BC induced vibration is observed in the FE model. This motion, which is shown in Figure 17, is qualitatively observed to be mostly comprised of rigid-body motion components (translation and rotation), with some deformational vibration components.

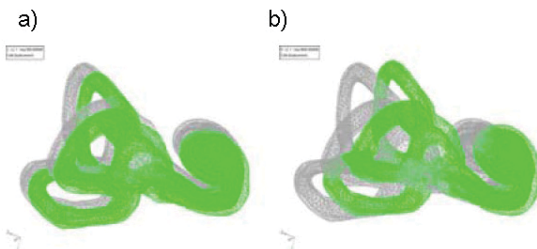


Figure 17. Snapshot images of the vibration of the cochlea realistically attached to the skull subjected to BC excitation at (a) 500 Hz and (b) 4000 Hz. Green image is the displaced state, and the gray image is the undeformed state. (Not to scale)

5.2 Sound transmission path analysis

The paths of sound transmission from the noise source via the flesh/plug system to the TM are investigated. Figure 18 below shows a conceptual diagram of the sound transmission paths for an ear fitted with an earplug. The external sound pressure acts on both the surface of the circumaural flesh, A, and the back surface of the plug, B, which in turn vibrates the tied plug/flesh system. The displacement degrees of

freedom of the flesh and plug are mutually coupled, as indicated by the bi-directional arrow C in the diagram. The occluded acoustic pressure in the ear canal is generated through surface vibrations of the tip of the earplug (E), as well as the flesh (D) lining the ear-canal wall.

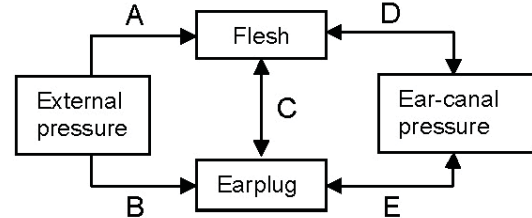


Figure 18. The transmission path diagram

The relative strength of each of these transmission paths can be analyzed in the FE model by selectively eliminating the coupling interfaces, which are labeled in Figure 18 as components A through E, and examining the resulting impact on TM pressure. Figure 19 below shows changes in resulting TM pressure due to different coupling conditions. The 0 dB line is the baseline case with all coupling interfaces intact. The earplug insertion depth used in these analyses is representative of a typical deep-insert earplug, whose tip is located near the bony canal entrance.

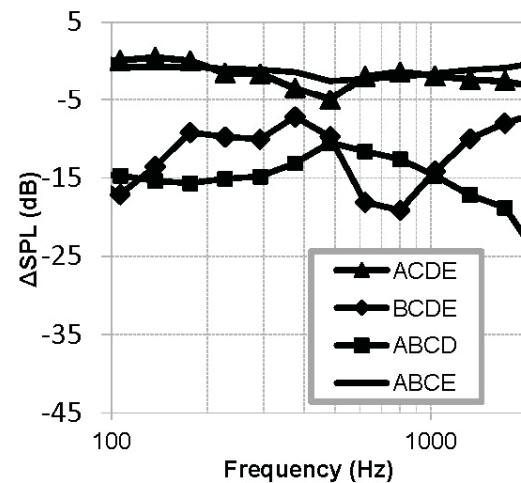


Figure 19. Changes in TM acoustic pressure (Δ SPL) by selectively de-activating coupling interfaces. The dB changes are relative to the baseline case where all interfaces are active. For example, ACDE stands for the case where the interface B is deactivated while other interfaces are intact.

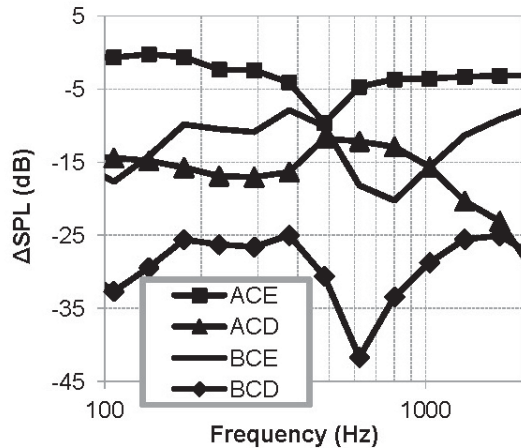


Figure 20. Δ SPL for different pathways

Figure 19 shows that effectively eliminating interface B (case ACDE), does not result in significant reductions in TM pressure as compared with eliminating interface A (case BCDE). This suggests that the relative contribution of the sound-to-flesh coupling is larger than that of the sound-to-plug coupling. This is expected considering that the area of the flesh where the incident sound acts upon is significantly larger than that of the plug. From the ABCD and ABCE results in Figure 19, it can be deduced that the main source generating TM pressure can be attributed to the vibrating surface of the earplug tip rather than that of the ear canal flesh. Again, this is expected considering that the deep-insert earplug constrains large portions of the thicker ear canal flesh, and terminates near the body canal entrance of the bony ear canal where thinner flesh layers have minimal contribution on occluded air pressure generation. However, it should be noted that the relative contribution of the flesh wall to the TM pressure would increase if the plug insertion depth is reduced, which exposes more of the thicker vibrating flesh wall. Further analysis is needed in conjunction with the insertion depth analysis described in the previous section. Figure 20 shows cases where two coupling interfaces (one outer and one inner) are deactivated. The results indicate that the strongest path is ACE (outer pressure $\rightarrow$ flesh outer surface $\rightarrow$ plug $\rightarrow$ TM pressure) while the

weakest is BCD (outer pressure $\rightarrow$ plug outer surface $\rightarrow$ canal wall flesh $\rightarrow$ TM pressure).

6 CONCLUSIONS

A general framework for developing detailed biomedical numerical models for AC and BC pathways, including ear insert HPD in high noise environments is presented. Starting with *in vivo* CT scanning to obtain an individual's ear canal, then using microCT scanning to obtain middle ear and cochlea geometries of generalized cadaver subjects, geometrical components of the biomedical FE are developed. The material properties of the components are initially acquired from literature information. Using several responses taken from published data for each component, initial material properties are tuned iteratively to exhibit consistent response. It is shown that the FE model may be simplified based on the frequency of interest. For example, the skull has a natural frequency of 1 kHz, therefore simplification can be obtained by imposing proper boundary conditions up to 1kHz instead of including the full skull in the FE model, which is effectively inactive in this frequency region. The completed and validated FE model is used to answer many challenging questions plaguing the hearing protection community. In this study, the BC limit and noise transmission pathways of flesh/plug systems are investigated. It is shown that there exists 5dB error in BC limit calculations of the FE model as compared to published experimental data. Considering the variability of human anatomical measurements and material properties, the obtained results are considered to be in agreement with literature results. Obtaining validated FE models enables the future development of advanced hearing protection systems, as well as novel hearing protection evaluation systems that can provide experimentally correlated psychoacoustic noise attenuation [24]. Explained FE model methodology and techniques may be further used to simulate the effects of blast related impulse noise on human auditory mechanisms and the brain.

7 ACKNOWLEDGEMENT

Research was supported by the Air Force Office of Scientific Research (AFOSR) under STTR funding (Contract NO: FA9550-06-C-0039) and grant DC 007910 from the NIDCD of NIH. Continuing work is currently being funded by NAVAIR SBIR funding (Contract No. N68335-10-C-0430).

REFERENCES

- 1 Berger, Elliott H., Kieper, Ronald W., and Gauger, Dan. Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways. *J. Acoust. Soc. Am.*, 114, 4 (October 2003), 1955-1967.
- 2 Koike, T., Wada, H., and Kobayashi, T. Modeling the human middle ear using the finite-element method. *J. Acoust. Soc. Am.*, 111, 3 (2002), 1306-1317.
- 3 Gan, R. Z., Feng, B., and Sun, Q. Three-dimensional Finite Element Modeling of Human Ear for Sound Transmission. *Annals of Biomedical Engineering*, 32, 6 (2004), 847-859.
- 4 Steele, C.R. and Taber, L.A. Comparison of WKB calculations and experimental results for three-dimensional cochlear models. *J. Acoust. Soc. Am.*, 65 (1979), 1007-1018.
- 5 Pathasarathi, A.A., Grosh, K., and Nuttall, A.L. Three-dimensional numerical modeling for global cochlear dynamics. *J. Acoust. Soc. Am.*, 107, 1 (2000), 474-485.
- 6 Stenfelt, Stefan. Acoustic and physiologic aspects of bone conduction hearing. *Advances in oto-rhino-laryngology*, 71 (2011), 10-21.
- 7 von Békésy, G. Vibration of the head in a sound field and its role in hearing by bone conduction. *J. Acoust. Soc. Am.*, 20, 6 (1948), 749-760.
- 8 Stenfelt, S. Transmission properties of bone conducted sound: measurements in cadaver heads. *J. Acoust. Soc. Am.*, 118, 4 (2005), 2373-2391.
- 9 Stenfelt, S. Vibration characteristics of bone conducted sound in vitro. *J. Acoust. Soc. Am.*, 107, 1 (2000), 422-431.
- 10 Aibara, R., Welsh, J.T., Puria, S., and Goode, R.L. Human middle-ear sound transfer function and cochlear input impedance. *Hearing Restoration*, 152 (2001), 100-109.
- 11 Lee, C., Chen, P., Lee, W., Chen, J., and Liu, T. Three dimensional reconstruction and modeling of middle ear biomechanics by high-resolution computed tomography and finite element analysis. *Laryngoscope*, 116 (2006), 711-716.
- 12 Lee, C.F., Hsu, L.P., Chen, P.R., Chou, Y.F., Chen, J.H., and Liu, T.C. Biomechanical modeling and design optimization of cartilage myringoplasty using finite element analysis. *Audiology & Neuro-Otology*, 11, 6 (2006), 380-388.
- 13 Homma, K. *Advanced Finite Element Model Development for Characterization and Mitigation of Bone Conducted Sound Transmission in High Noise Environment*. STTR Phase II Contract Number FA9550-07-C-0088, 2010.
- 14 Gan, R.Z., Reeves, B.P., and Wang, X. Modeling of Sound Transmission from Ear Canal to Cochlea. *Annals of Biomedical Engineering*, 32 (2007), 847-859.
- 15 Homma, K., Du, Y., Shimizu, Y., and Puria, S. Ossicular resonance modes of the human middle ear for bone and air conduction. *J. Acoust. Soc. Am.*, 125 (2009), 968-979.
- 16 Merchant, S.N., Ravicz, M.E., and Rosowski, J.J. Acoustic input impedance of the stapes and cochlea in human temporal bones (1996).
- 17 Puria, S., Peake, W.T., and Rosowski, J.J. Sound-pressure measurements in the cochlea vestibule of human cadaver ears. *J. Acoust. Soc. Am.*, 101 (1997), 2754-2770.
- 18 Nakajima et al., H.H. Clinical investigation and mechanism of air-bone gaps in large vestibular aqueduct syndrome. *The Annals of Otolaryngology, Rhinology, and Laryngology*, 116, 7 (2007), 532-541.
- 19 Greenwood, D.D. A cochlear frequency-position function for several species-29 years later. *J. Acoust. Soc. Am.*, 87, 6 (1990), 2592-2605.
- 20 Stenfelt, S., Puria, S., Hato, N., and Goode, R.L. Basilar membrane and osseous spiral lamina motion in human cadavers with air and bone conduction stimuli. *Hearing Restoration*, 181 (2003), 131-143.
- 21 Gunderson, T., Skarstein, O., and Sikkeland, T. A study of the vibration of the basilar membrane in human temporal bone preparations by the use of the Mössbauer effect. *Acta Otolaryngol.*, 86 (1978), 225-232.
- 22 Zwislocki, J. In search of the bone-conduction threshold in a free sound field. *J. Acoust. Soc. Am.*, 29, 7 (1957), 795-804.
- 23 Reinholdt, S., Stenfelt, S., and Håkansson, B. Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations. *J. Acoust. Soc. Am.*, 121, 3 (2007), 1576-1587.
- 24 Schank, K., *Development of Advanced Hearing Protection Evaluation System*. SBIR, Contract No. W911QX-10-C-0080.

2.3 Web-Based Predictive Analytics to Improve Patient Flow in the Emergency Department

Web-Based Predictive Analytics to Improve Patient Flow in the Emergency Department

David L. Buckler, M.S.
CACI International, Inc.
dbuckler@caci.com

The Emergency Department (ED) simulation project was established to demonstrate how requirements-driven analysis and process simulation can help improve the quality of patient care for the Veterans Health Administration's (VHA) Veterans Affairs Medical Centers (VAMC). This project developed a web-based simulation prototype of patient flow in EDs, validated the performance of the simulation against operational data, and documented IT requirements for the ED simulation.

In addition to enabling VHA to provide higher quality, more convenient and comfortable care for veterans and their families, the application will provide the VHA with the following benefits:

- Web-based user interface to manage key input parameters
- Improved analytical approach to managing ED processes
- Better understanding of the operational variances within EDs
- Improved resource management
- Proactive versus reactive decision making
- Well-defined management reporting requirements
- Adherence to Federal patient care regulation guidelines
- Integration of simulation with Emergency Department Information System (EDIS) v1.5 rollout

The simulation application, built using CACI's SIMPROCESS™ software, is scalable so it can be deployed across the entire VHA. The initial focus was on VAMC Emergency Departments, but eventually the simulation will be expanded to include all inpatient flow processes, including bed management and surgical scheduling.

Goals of the Project

VHA EDs received both short term and long term benefits from this project. In

the short term the project helped EDs standardize their real-time data collection. In the long term simulation will help EDs develop better processes, improved efficiency, and optimal patient flow. ED personnel can use the web-based interface to examine their own "what-if" scenarios such as reducing the total number of patients waiting more than 6 hours for admission, reduce the total time EDs spend on diversion, reducing missed patient opportunities, and increasing the percentage of discharges before noon.

Overview of Concept

VAMC personnel need accurate and timely data to make smart decisions regarding patient care and resource planning. Simulation is an effective way to look at a system's current and future state performance, however most VA personnel don't have the technical training to build and execute a simulation model of their business processes. The concept of this project is to put these analytical capabilities into the hands of a VA decision maker rather than a trained simulation analyst. The simulation tool enables decision makers to access standard desktop simulation models through a stand-alone, web-based interface allowing them to change key parameters, e.g., mix of patient acuity and analyze the impact of these changes.

The diagram below shows the integration of data analysis into ED patient care. As patients are treated in the ED, data is collected by EDIS and other computer systems. Most of the patient data is entered by ED personnel, including patient arrival time, patient acuity, physical symptoms, medical diagnosis, and patient discharge time.

All of this data can be analyzed to determine the effectiveness of an ED's business processes. This tool can import all patient data and generate real-time and future-state reports which allows for immediate and long-term planning. As shown below, this is a continuous cycle of improvement.

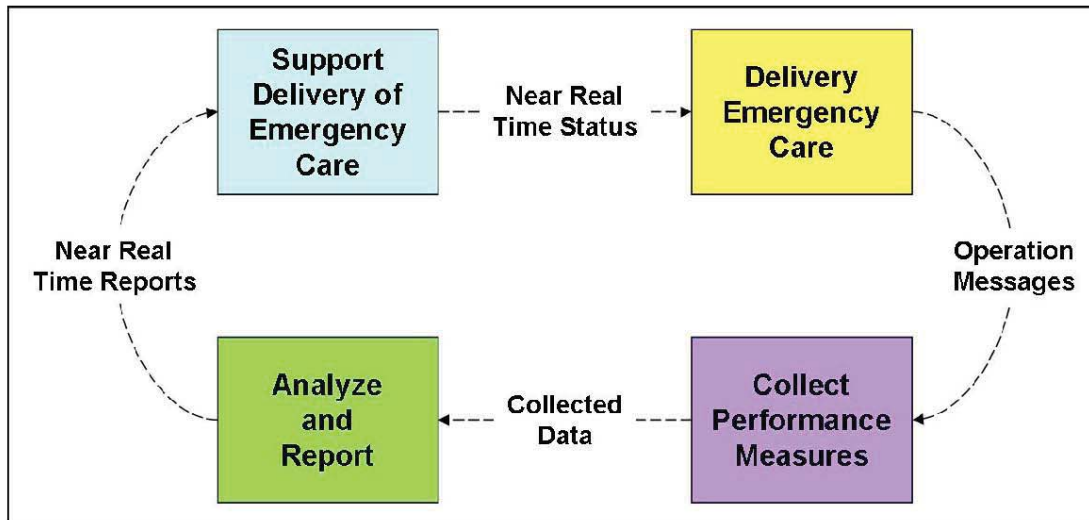


Figure 1 – Continuous Cycle of Improvement

The approach used in this project is intended to transform 'level 1' simulations to 'level 2' simulations and establish a framework for a real-time predictive analytics tool. The approach is as follows:

- **Step 1** – Build individual simulation models of several VAMC EDs. Personal interviews, site visits, and system data are used to build these models, which must be validated before moving to the next step. Apply stochastic variation to patient arrival and task completion to replicate the variability that exists in real-world settings.
- **Step 2** – Combine the individual models into one business process model.
- **Step 3** – Build a web-based interface and data repository that

connects to the combined business process model.

- **Step 4** – Run various "what-if" analyses on the models to examine the impacts of adding additional resources, inpatient beds, etc.
- **Step 5** – Build a template library that will contain the key tasks, resources, and activities associated with ED business processes. These templates can be accessed by any VAMC to build their own respective simulation.

Building the decision support application in SIMPROCESS creates a migration path to a 'level 3' solution, the most powerful application of simulation because it provides real-time predictive analytics and Business Activity Monitoring (BAM). By integrating the

simulation tool with current ED applications, administrators can monitor the status and performance of ED operations and make proactive, versus reactive, decisions.

The diagram below shows the planned interactions between the various inpatient flow management information systems. The EDIS, Bed Management,

and Surgical Scheduling applications share patient data with each other to minimize the required data entry and maximize the efficiency of patient flow. The data from these applications can also be sent to the SIMPROCESS ED model, and through the web-based user interface ED employees and managers can use this real-time data for current-state and future-state process analysis.

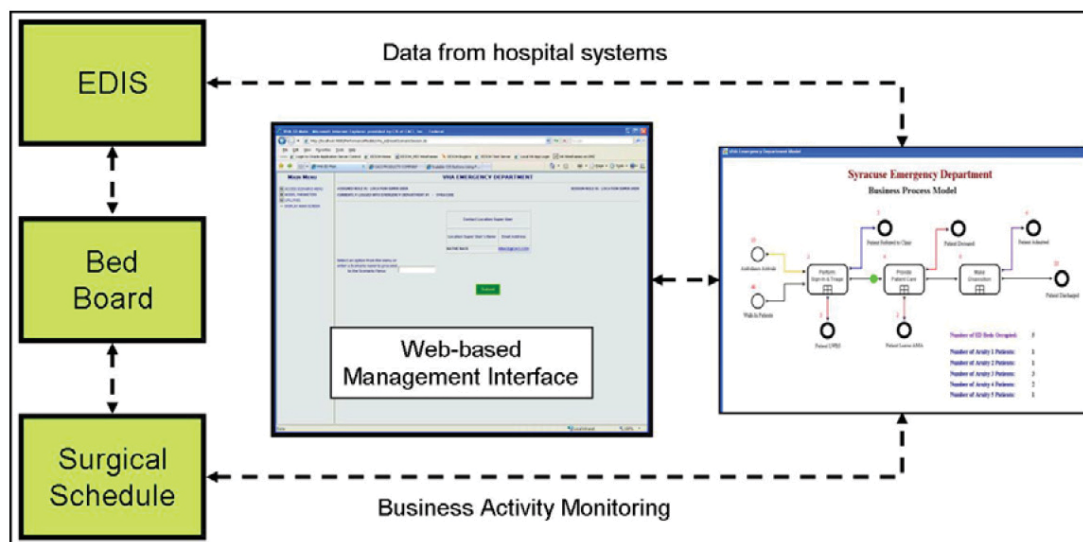


Figure 2 – Visualization of System Communications

Project Activities

The project began with a business process study of the Syracuse VAMC ED. This study included personal interviews with ED staff members, personal observations of the ED, and analysis of ED Tracker data, which provided the foundation for the Syracuse ED simulation model. This model included the patient profile, resource utilization, and business rules of the Syracuse ED. An initial draft of the model was reviewed by Syracuse administrators and the appropriate changes were made to validate the model. The model was considered valid when the simulation results for patient

arrivals, patient cycle times, and resource utilizations were statistically similar to real-life data.

The second step of the project was to conduct a similar business process study of the Lexington VAMC ED. Like Syracuse, this study consisted of personal interviews with ED staff members, personal observations of the ED, and analysis of their ED Tracker data. Based on this analysis, a second simulation model was built. This model not only included the patient profile, resource utilization, and business rules of the Lexington ED but showed the variations in patient care between the two ED sites. An initial draft of the model was reviewed by Lexington

administrators and it was validated in much the same manner as the Syracuse model.

Once both of these models were complete, a front end application was developed that provided access to each model via an Internet Explorer web browser. The process models themselves were transferred to a database and hosted on a server, which connects to the web-based user interface. This transition allows VA users to build scenarios, change a few key parameters, run simulations, and conduct future-state “what if” analysis.

St. Louis was the third VAMC ED chosen for the prototype. After a site visit to St. Louis, a St. Louis business process model was built, validated, and integrated into the existing web-based simulation prototype. This completed the activities associated with the first phase of the project.

The second phase of the project will focus on making the simulation prototype both scalable and flexible so it has utility for every VAMC ED. The

initial prototype will serve as the foundation of a template library of process, activities, and resources. This library will allow any ED to build its own business process model and execute simulations. The user will be able to change any model parameter and the tool will feature a more robust visualization and reporting capability.

Simulation Model

Figure 3 shows the top level of the Syracuse ED simulation model. The model follows patients from the time they enter the ED until the time of their disposition, highlighting the 3 main processes associated with patient care: “Perform Sign-In and Triage”, “Provide Patient Care”, and “Make Disposition”. Patients can arrive via ambulance or through the main door. The model captures all of the different ways a patient leaves the ED: discharged by the physician, admitted to the inpatient facility, against medical advice, referred to a clinic, deceased, or without being seen.

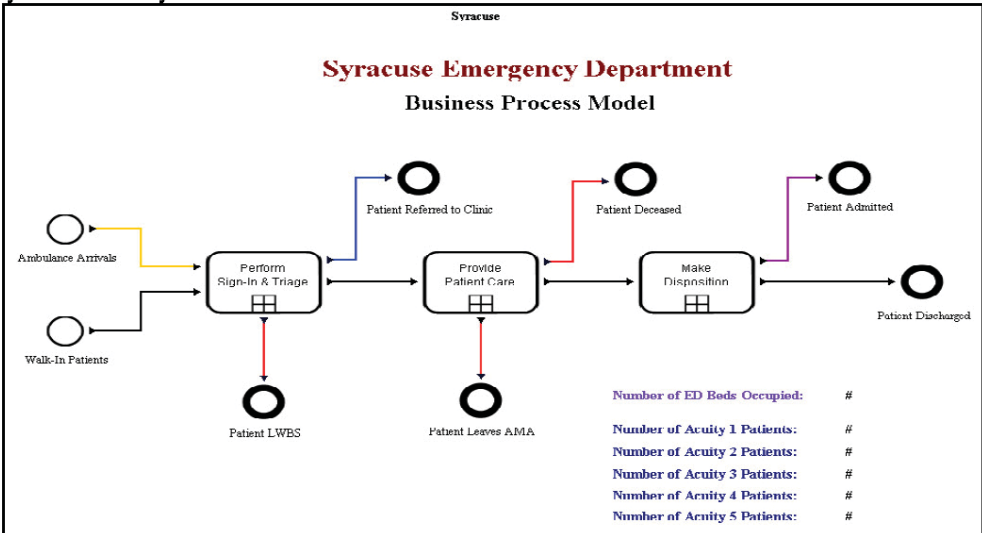


Figure 3 – Syracuse VAMC ED Business Process Model

The Lexington and St. Louis ED business process models are broken out into the same general structure, however, each site contains variations in patient population, resource utilization, and business practices. Combining

these models into a template library allows other ED sites to build their own unique simulation model from a group of prebuilt constructs which reduces the time to generate results.

Web-based User Interface:

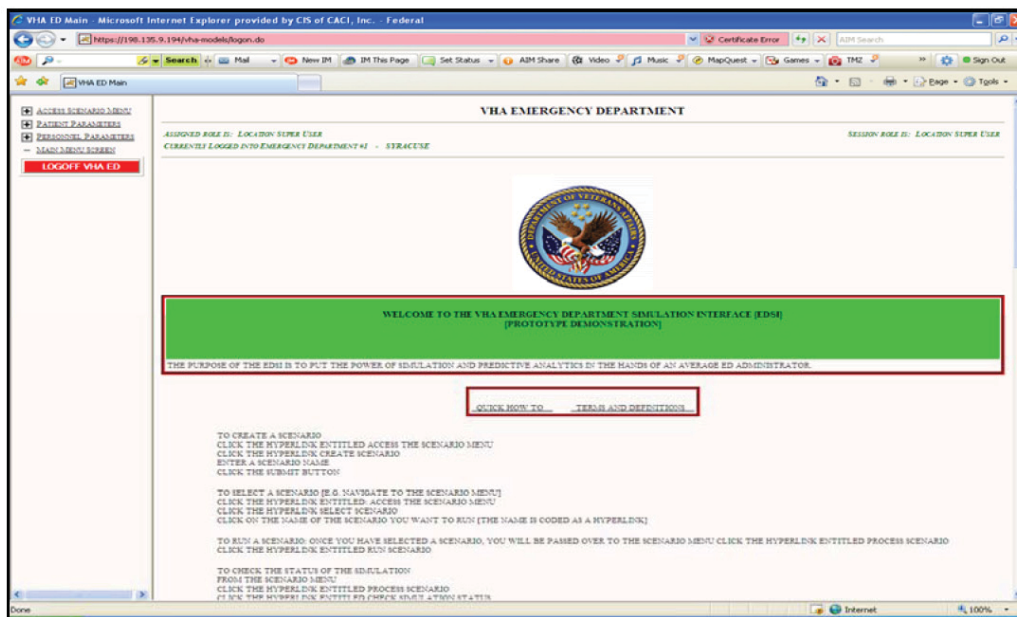


Figure 4 - Main user interface screen

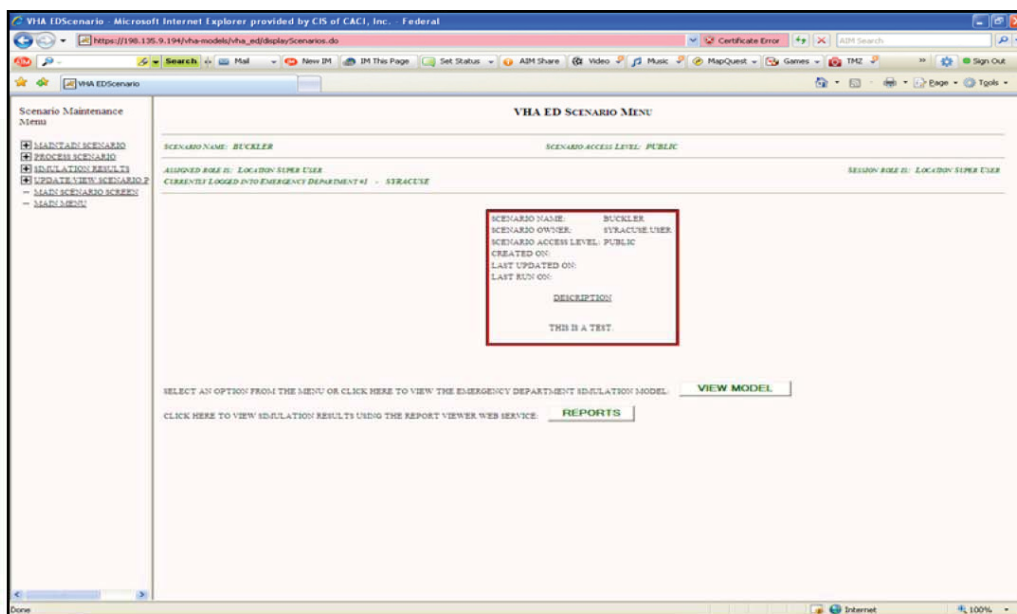


Figure 5 – Scenario creation screen

Lessons Learned

Building an ED simulation prototype produced a number of “Lessons Learned” that can be applied to future prototype efforts. The lessons learned encompass the technical aspects of the simulation, the challenges of working with VA patient data, and the variations of ED business processes.

Process Variations:

Three VAMC ED sites were included in the simulation prototype: Syracuse, Lexington, and St. Louis. Based on our site visits, we concluded that the foundation of ED patient care consisted of three treatment stages: Sign-in and Triage, Provide Patient Care, and Make Disposition. However these sites revealed that there are significant process differences between each site, and that reinforced the need for a flexible tool that can be adjusted to reflect the business processes at any ED facility. Some of these process differences include:

- The existence of an Urgent Care center
- Sign-in / Triage processes
- Resource shifts are scheduled differently at various sites.
- Location of lab and testing equipment greatly impacts a patient's cycle time.
- When a patient is waiting for a bed in the inpatient facility, which should get charged that delay time, the ED or the hospital?

Data Issues:

The long-range goal is to connect the ED simulation prototype to other patient

flow prototype models, primarily a Bed Management and Surgical Scheduling application. Connecting to these other models may require data elements that are not currently collected or under consideration. However the ED prototype is flexible enough to add in new data and demonstrate the expected impacts. Other data observations include:

- When working with VA patient data it is imperative that personal identifiers be removed, such as social security numbers, etc.
- Each site can provide information concerning resources and resource scheduling within their respective ED. Users should be able to change all of these parameters because resource management is a significant concern to hospital administrators.
- EDs should collect more accurate data regarding patient tests and consults. Better data would allow the simulation models to more accurately calculate the average number of tests ordered (per acuity level), the average number of consults ordered (per acuity level), and the travel time to and from lab equipment. This could vary widely from one ED facility to another and cause unwanted patient delays.
- The following data elements that should be collected by EDIS or clarified by the VHA:
 - EDIS should standardize disposition codes across all EDs.
 - The number of tests and consults ordered per patient should be entered into EDIS.

- The exact patient “check-in” time to the ED should be standardized. Is it when a patient walks through the door, when a patient signs in, or when a patient is seen by a health care provider first sees a patient?

Conclusions

Each site agreed that a simulation tool would be very useful in helping them manage their ED processes. The ED simulation prototype provides a test environment that can assist in evaluating patient flow improvements and it establishes a quantitative basis for decision-making that allows comparison among different types of business processes, resource staffing, and patient care options. In the near-term, the prototype provides a tool for continuous improvement, and in the long run it supports documentation and sharing of best practices among VAMC facilities.

The unique composition of the VA's patient population changes the way users should analyze their business processes. Since many ED patients return at some point, performance metrics should be understood to reflect the efficiency of an ED, not necessarily the quality of care provided. Trying to measure overall patient satisfaction and quality of patient care is very difficult and requires different analytical techniques such as conducting patient surveys.

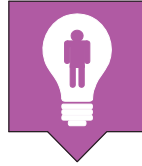
The ED simulation prototype allows users from the Syracuse, Lexington, and St. Louis EDs to study their respective business processes. They can change

a few key parameters and conduct “what-if” future-state analysis. Although the ED simulation prototype produces accurate results when compared to real ED data, more testing and consult data is required to gain full confidence in the results. Bed requirements are needed when admitting patients to the inpatient facility. The current prototype does not model to this level of detail and it will need to be adjusted before it can connect to the overall flow management model.

This simulation project proved that despite several real-life data shortcomings an effective tool could be developed from personal observations, smart statistical assumptions, and available process data. If the VA starts collecting additional data, the simulation can be refined to provide even better predictive analytics.

3.0 HUMAN DIMENSION TRACK

Track Chair – Dr. Kara Latorella, NASA Langley Research Center



3.1	Using M&S to Improve Human Decision Making and Achieve Effective Problem Solving in an International Environment.....	162
3.2	Developing Culture-Adaptive Competency through Experiences with Expressive Avatars	166
3.3	Using Agent Based Modeling (ABM) to Develop Cultural Interaction Simulations	175
3.4	Cognitive Network Modeling as a Basis for Characterizing Human Communication Dynamics and Belief Contagion in Technology Adoption.....	181
3.5	A Systematic Approach for Real-Time Operator Functional State Assessment	190
3.6	Engagement Assessment Using EEG signals	200
3.7	Modern Warfare: An M&S Examination of the Dynamic Impact of Warlords and Insurgents on State Stability	208
3.8	Using Random Forest Models to Predict Organizational Violence.....	216
3.9	A Model of Rapid Radicalization Behavior Using Agent-Based Modeling and Quorum Sensing.....	223
3.10	The Human Dimension of Closing the Training Gap for Fifth-Generation Fighters	231
3.11	Physiological Based Simulator Fidelity Design Guidance	238
3.12	Predicting the Impacts of Intravehicular Displays on Driving Performance with Human Performance Modeling.....	248
3.13	Understanding the Impact of User Frustration Intensities on Task Performance Using the OCC Theory of Emotions	254
3.14	Automatic Facial Expression Recognition and Operator Functional State	258
3.15	A Human View Model for Socio-Technical Interactions	268
3.16	Understanding Group/Party Affiliation Using Social Networks And Agent-Based Modeling.....	275
3.17	The National Operational Environment Model	281

3.18	Exploring the Use of Computer Simulations in Unraveling Research and Development Governance Problems.....	289
3.19	An Agent-Based Model Of New Venture Creation: Conceptual Design for Simulating Entrepreneurship	299

3.1 Using M&S to Improve Human Decision Making and Achieve Effective Problem Solving in an International Environment

Using M&S to Improve Human Decision Making and Achieve Effective Problem Solving in an International Environment

Vanessa L. Christie & David J. Landess
Prevailance, Inc.

vanessa.christie@prevailance.com david.landess@prevailance.com

Abstract. In the international arena, decision makers are often swayed away from fact-based analysis by their own individual cultural and political bias. Modeling and Simulation-based training can raise awareness of individual predisposition and improve the quality of decision making by focusing solely on fact vice perception. This improved decision making methodology will support the multi-national collaborative efforts of military and civilian leaders to solve challenges more effectively. The intent of this experimental research is to create a framework that allows decision makers to "come to the table" with the latest and most significant facts necessary to determine an appropriate solution for any given contingency.

Introduction

Imagine yourself sitting down in a meeting surrounded by individuals from many nations, none your own. Who do you gravitate toward? Who do you seek as an early ally? Why? These are some of the questions that we plan to address in our research. Modeling & Simulation (M&S) based training can create a realistic and relevant forum to explore our own predispositions and biases. Every individual has specific factors that influence their decisions and the multi-national collaborative environment often presents multiple factors that increase the likelihood of poor decision making. Our premise is that effective decisions are those based on facts, not perceptions. Our intent is to identify the role of individual partiality in decision making and clearly define the factors that dissuade each individual from the facts necessary to make an effective decision. Furthermore, we intend to present a framework that facilitates fact-based decision making in an international environment.

Fact-based decision making requires a focus on assessing the capabilities that each nation might contribute to solve the challenges facing the group. However, that capabilities assessment is often influenced by the language bias of the individuals conducting the assessment. According to research conducted by psychologists at

Harvard, the University of Virginia and the University of Washington, "A growing number of studies show a link between hidden biases and actual behavior. In other words, hidden biases can reveal themselves in action, especially when a person's efforts to control behavior consciously flags under stress, distraction, relaxation or competition." [1] For example, if you sit down in a group of people from all over the world and you appear to be the only one speaking English, you generally reduce the amount of input that you add to the conversation because it is uncomfortable to not be understood. Similarly, if you see there is one more English speaker, you will be inclined to converse with just that individual. This is where our bias lies. Most people affiliate with others who speak a common language. It makes sense. You want to communicate with those who can communicate in return. However, this bias often prevents the group from discovering the availability of the very capability that will contribute most to the challenge, leading to a poor, non-fact-based decision. Culture is another factor that influences individuals and contributes to the decisions made by collaborative teams in an international environment. A study conducted by a Swedish university used a M&S based methodology to investigate the various characteristics that influence decision making and found that cultural elements, such as a propensity to delegate

vice take charge in the face of an emergency situation, played a significant role in how teams from different nations interacted and how that influenced group decisions. "The disparity in ways to be diverse is likely to be the rule rather than the exception...Managers and leaders who assemble multinational teams should be prepared for this disparity in modes of cultural diversity" [2]. Unfortunately, the study did not attempt to measure the effectiveness of the decisions, but with some modifications to the M&S methodology, that determination could be made.

The aforementioned biases in language and culture have manifested themselves in the propensity for the United States to consistently partner with the United Kingdom, Canada, Australia and New Zealand in both civilian and military endeavors. The U.S. has closed computer networks to only work with these four nations and the U.S. moves through most conflicts in a strict partnership that is mutually beneficial with these nations. Although this relationship has been effective, it is possible that the U.S. may have had other foreign assets available for coalition operations with direct applicability to a given geographic region or for a specific conflict. It appears that the close alignment in language and culture between all five nations played a part in our collective history and the decisions made therein. That is not to say that with a capabilities assessment our leaders would not have made the same decision, but it does indicate a definitive partiality that facilitates interaction within the circle of five nations and creates barriers for entry with other nations.

In order to further examine the factors that influence our individual biases in a collaborative situation, we have to look at the individual elements of our personalities that impact a team setting. In addition to language and culture, the larger set of biasing factors includes politics, financial status, physical appearance, personal mannerisms, marital status, age and race.

In general, one factor alone will not lead an individual away from fact-based decision making, but with multiple factors in play; it is likely we will see a breakdown in effective problem solving. We stray from the facts and rely on what we see or "know" from our upbringing and/or life experiences.

Even though we can identify the individual factors that contribute to bias or predisposition, the international arena creates additional factors for consideration. It forces each individual to look at differences in culture and factor them into a decision. However, factoring in culture does not necessarily make the solution set easier, in fact, it complicates it to the point where the requisite facts are harder to identify and quantify. For example, imagine yourself as one of five commercial manufacturing plant leaders in the world. There are representatives from four other countries and two translators in the room. Although there are clear and concrete facts that need to factor into your decision to move a portion of your plant to these nations, you are strangely distracted by the dialect of one participant and the professional attire of another. It could be assessed to be a lack of focus due to these unique differences between yourself and the other participants, but more significantly, it may be tied to the same elements of your own predisposition. If you knew that you would be distracted coming in the door, then you could have quickly addressed the valid points for your argument to include the manufacturing plant specifics of available labor force, price per square foot, environmental constraints, etc.

So, the question then becomes how do we get back to the facts...and push all of this other information out of our minds. By modeling this type of situation with various players in dynamic environments, individuals will quickly realize the factors that most affect their decision making. The next step is to eliminate this partiality and get back to the facts.

Discussion

Before discussing an approach to getting back to fact-based decision making, it is important to bring up two other elements that complicate collaborative efforts in an international environment - particularly those conducted during military operations. Those elements are time and stress. During most military contingency operations, it can be extremely challenging to get all of the facts before a decision needs to be made. For example, if a natural disaster has left hundreds of people helpless and stranded, the need to act far outweighs the time required to obtain all of the facts. Similarly, if we look to our military, we can identify numerous situations where lives are at stake and a rapid decision needs to be made. In that regard, we need to use M&S to help each leader to define his, or her, own elements of predisposition and find a way to acquire facts in an expeditious manner. Military leaders often face life or death consequences based on their decisions; the stress of rapid response must also be addressed in order to prevent it from enhancing our natural biases that run counter to our desire for fact-based decisions.

Research conducted by the University of Pennsylvania concludes that the need for human behavior models is nowhere more apparent than in the military modeling and simulation community ...to satisfy a wide and expanding range of scenario concerns. Their interest goes beyond mission-oriented military behaviors, to also include simulations of the effects that an array of alternative diplomatic, intelligence, military, and economic (DIME) actions might have upon the political, military, economic, social, informational (psyops), and infrastructure (PMESII) dimensions of a foreign region" [3]. Only by expanding the current research conducted to date on DIME and PMESII and creating an M&S environment reflective of multinational collaborative teams will we be able to accurately assess an international environment and arm our decision makers

with the fact-based information they need to make sound decisions.

"Operations Iraqi Freedom and Enduring Freedom have demonstrated the need for rapid change in tactics, techniques, and procedures and our overall approach to campaigning. They have proven that the more complex the contemporary operational environment, the more the body of professional military knowledge must remain in a state of purposeful instability [4]." With this purposeful instability comes the need for a reduction in bias and an open mind for all possible options. An international environment adds to the overall complexity of any given campaign and forces an understanding above and beyond decision making in a single nation setting. With that being said, we can recreate actual environments from specific military conflicts using M&S to better prepare our military leaders. We can take direct examples from military leaders who made what they believed to be the best decision at the time and analyze it through interactive training.

The M&S environment could fall into a "gaming" category where different role players represent different nations and each nation's intent in the scenario; or the M&S solution could be as simple as a single role for the participant and all other players are programmed as they were during that day in history. We can evaluate the decision of the M&S trained individual and compare it to the decision of the actual military leader during the historic event. This comparison will not only force leaders to learn from historic lessons, but also grow awareness for the pitfalls of DIME and PMESII.

This same concept can apply to the civilian sector. If we take the actual experiences of civilian business leaders and recreate each scenario in an M&S world - civilian business leaders can train in this world. Not only will they have the ability to evaluate themselves, but also reflect on the decision made by the business leader with the same information presented. This will help them to identify the information that is truly needed, that which causes distraction

and the combination of the above that works for each civilian leader to truly make a fact-based decision.

Conclusion

In summary, M&S based training is a viable mechanism for decision makers in both the civilian and military sectors to overcome their biases and become better at focusing on relevant facts. In order for our leaders and decision makers to quickly capture a partner nation and/or organization's capacity and capability, we need to incorporate known predisposition and requisite capabilities assessment with M&S scenarios. We believe this includes capabilities assessment injected directly into the current decision making portion of any given process. In order to remain as inclusive as possible for a broad spectrum of challenges, we plan to include international organizations (IO), regional governmental organizations (GO) and non-governmental organizations (NGO) in our study as there is a tremendous difference in the military versus civilian mindset and model therein despite the fact that the capabilities of each enhance the effectiveness of the other. We understand that it is impossible to obliterate a bias, but it is possible, from an academic and M&S standpoint, to objectionably assess a situation, look at all of the pertinent factors and facilitate an effective decision to align with the highest value assets in an international environment.

References:

- [1] Test Yourself for Hidden Bias. 2011. Retrieved July 15, 2011 from <http://www.tolerance.org/activity/test-yourself-hidden-bias>
- [2] Smith, Kip. Empirical studies and an explanatory model of cultural differences in goal setting, task allocation and communication. Linkopings Universitet – Sweden, 2010.
- [3] Silverman, Barry; Bharathy, Gnana; O'Brien, Kevin; Cornwell, Jason. *Behavior Models for Agents in Simulators and Games*. (2006). Retrieved July 15, 2011 from www.seas.upenn.edu/~barryg/PRESENCEpt2.doc
- [4] Paparone, Christopher and Reed, George, The Reflective Military Practitioner: How Military Professionals Think in Action, *Military Review*, Mar-Apr 2008. Retrieved August 14, 2011 from http://www.au.af.mil/au/awc/awcgate/milreview/paparone_mar08.pdf

3.2 Developing Culture-Adaptive Competency through Experiences with Expressive Avatars

Developing Culture-Adaptive Competency through Experiences with Expressive Avatars

Daniel S. Silverglate
DanS@vcom3d.com

Edward M. Sims, PhD
EdS@Vcom3D.com

Vcom3D, Inc.

Gerald Glover, PhD
jgloverocd@yahoo.com
Hawaii Pacific University

Harris Friedman, PhD
harrisfriedman@hotmail.com
University of Florida

Abstract. Modern Warfighters often find themselves in a variety of non-combat roles such as negotiator, peacekeeper, reconstruction, and disaster relief. They are expected to perform these roles within a culture alien to their own. Each individual they encounter brings their own set of values to the interaction that must be understood and reconciled. To navigate the human terrain of these complex interactions, the Warfighter must not only consider the specifics of the target culture, but also identify the stakeholders, recognize the influencing cultural dimensions, and adapt to the situation to achieve the best possible outcome. Vcom3D is using game-based scenarios to develop culturally adaptive competency. The avatars that represent the stakeholders must be able to portray culturally accurate behavior, display complex emotion, and communicate through verbal and non-verbal cues. This paper will discuss the use of emerging game technologies to better simulate human behavior in cross-cultural dilemmas.

Nomenclature: culture, adaptive, values, cultural values dimensions, dilemmas, virtual humans, non-verbal communications

1.0 INTRODUCTION

This paper addresses a technology and methodology for the development and delivery of serious games for instruction in cross-cultural adaptability. The Cross-cultural Adaptivity Trainer (CCAT) system has been developed over the course of several projects, including cultural training for the US Navy in the Pacific, the Royal Australian Air Force (RAAF), the Air Force Negotiation Center of Excellence (NCE), and the Joint Warfare Center (JWFC).

Whereas the present paper describes the technology we have developed to create and deliver Cross-cultural Adaptability

Training (CCAT), a companion paper by the same authors [1] focuses on the underlying theory and pedagogical design.

2.0 CULTURALLY ADAPTIVE COMPETENCY

Contemporary Warfighters are expected to conduct operations and successful missions in many areas of the world. The global community includes a wide variety of socio-cultural contexts, each one with unique challenges for cross-cultural communication and adaptive competency. It is impossible to learn everything about all cultures in the world, or even everything about one culture. Yet we can observe that some people are able to effectively operate in a wide variety

of socio-cultural contexts, even when they do not have prior knowledge or experience of the local culture.

This ability to adapt to a variety of socio-cultural contexts is primarily learned, and is based on the development of adaptive predispositions for behaving effectively in a global community. Specifically, being culturally adaptive is based on predispositions in which a person is able to: 1) recognize and respect cultural differences; 2) make sense of signals and cues in a culturally diverse situation; 3) create synergy from diversity and resolve cultural dilemmas; and 4) view the socio-cultural encounter from a holistic perspective.

As shown in Figure 1, culturally adaptive competency requires that a person be able to recognize, respect, resolve, and reconcile cultural differences in socio-cultural encounters, even in socio-cultural contexts in which the person has no prior knowledge nor experience.

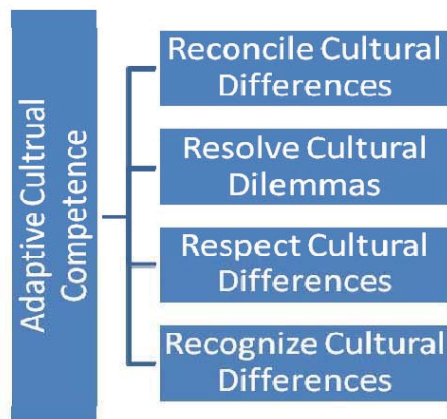


Figure 1. Adaptive Cultural Competence

A framework for understanding and making sense of cultural encounters is the underpinning for developing CCAT. Culture has many levels of understanding and abstraction, including behaviors, beliefs, and values. Values are ways that cultures provide meaning to human behavior and beliefs, and values represent what is important to people. Human behaviors and

beliefs can be seen as falling within the extremes of opposite positions on values dimensions. These values dimensions are not stereotypes, but represent dynamic tensions between opposites, such as individual versus collective orientations.

Cross-cultural dilemmas occur when actors hold seemingly opposing values related to a situation, issue, negotiation, or other aspect of a cross-cultural encounter. Cross-cultural dilemmas can be understood and effectively navigated when actors understand their own cultural values in relation to the values of other actors, and know how to resolve or reconcile important differences. Resolve refers to obtaining a workable compromise, whereas reconcile refers to obtaining a higher-order solution that exceeds a compromise and involves creating an innovative solution to a dilemma. Adapting cross-culturally is related to being able to appropriately resolve and reconcile cultural dilemmas through adaptive decision making. We use the concepts of cultural values dimensions, cultural dilemmas, and adaptive decision making to build the CCAT approach, which provides learners with opportunities to experience and respond to cultural dilemmas through recognizing and respecting different value dimensions and resolving or reconciling them through adaptive decision making.

3.0 OVERVIEW OF CCAT

CCAT focuses on developing the learners' cross-cultural adaptive predispositions by providing the learner with a working knowledge of culture through an easily useable framework for making culture operational, and opportunities to engage with scenarios and simulated experiences for applying this knowledge in a variety of cross-cultural encounters.

While video recordings of live role-playing scenarios could be an option for CCAT, we are developing the training using interactive

3D serious game technology. This approach provides several advantages:

- Scenarios and simulations can be created and edited rapidly without access to live actors.
- The technology supports naturalistic interactions, in which computer-generated characters take initiative, in contrast to branching video.
- Advances in virtual human technology support the inclusion of subtle nonverbal cues.

CCAT is currently under development and will be available from a Web site for universal access, as well as on mobile tablets and smartphones. Training will be delivered by avatar-based scenarios and simulations involving cultural dilemmas through which actors holding divergent cultural value orientations will need to navigate. Learners will view and respond to these in virtual cross-cultural encounters. In order to provide the variation and repetition required to support recognition of cultural values from context [2], we envision the development of 35 scenarios and 5 simulations to provide cross-cultural experiences in a globally-appropriate collection of cultural encounters. In a relatively brief time period, learners are expected to gain cultural knowledge and experience typically available only to persons who have traveled the world extensively for years.

4.0 TECHNOLOGY REQUIREMENTS

Experiential training for cross-cultural adaptability poses new requirements for serious games. Specifically, the learner needs to be exposed to simulated situations in which they must recognize subtle variations in behavior that are indicative of higher-level differences in cultural beliefs and values. The learner must then apply culture-general knowledge and metacognitive skills to form a mental model of the different cultural values influencing the situation.

We have identified the following technical requirements as key to the successful development of CCAT training:

- Computer-generated characters must be highly expressive, with nonverbal behaviors indicative of emotional response and focus of attention.
- The actors' behaviors must be highly dynamic, showing subtle variations in response to learners' actions.
- The scenarios and simulations should use authentic, relevant storylines to engage the student and keep them motivated.
- The technology should be capable of being adapted to emerging gaming platforms, including mobile tablets, smartphones and virtual worlds.

5.0 EXPRESSIVE AVATARS

Zbylut and Metcalf [3], in an ARI study, have indicated that skill at interpreting nonverbal behavior is one of the most frequent and important communication skills that advisors reported using while deployed. Effective training of culture-adaptive competency requires computer-generated characters capable of reproducing complex nonverbal behaviors such as body gestures, facial expressions, and eye movements.

Body gestures may be either universally recognizable body language or deliberate gestures unique to some culture. We have created a system to allow cultural experts to build their own gestures utilizing Inverse Kinematics (IK). As opposed to Forward Kinematics (FK), where the 3D character is posed by direct manipulation of each joint, Inverse Kinematics allows the hands or fingers to be placed relative to a set of focal sites on the body [4, 5]. A complex set of algorithms then determine the joint angles necessary to achieve this result. The algorithms take into consideration the stresses and limitations of each joint so that pose emulates proper human motion, but

can still be adjusted to ensure the gesture conveys the intended meaning. Since gestures created in this manner are target based, they can be applied to characters with different body types.

As reported by Paul Ekman [6], the perception of emotion through facial expressions is universal throughout all cultures. However, the rules for displaying emotion vary significantly according to the values of a culture [7]. For computer-generated characters to be perceived as having emotions, it is important to apply the fundamental aspects of how facial expressions are formed and their relationship to emotions. Ekman and his colleagues developed the Facial Animation Coding System (FACS) [8] as a mechanism to quantify the movement of facial muscles and relate them to the expressions they form.

According to FACS, facial expressions are the result of the activation of individual muscles or muscle groups beneath the skin. Similarly, the movement of the mesh of a seamless 3D model of the face is affected

by the articulation of elements known as bones. Each 3D bone affects some set of vertices in the mesh by some percentage called the weight. Each vertex of the face mesh may be influenced by more than one bone but the cumulative weights should total 100%.

We have developed a new bone-based rig for our 3D characters' faces. Each of 39 "bones" represents one of the FACS action units and by extension the muscle or muscle group of the human face. The bones can be animated to simulate the movement of their real world muscle counterparts. Figure 2 shows the process of creating a facial expression for an avatar. On the left is an image of the control rig which allows the animator to move control points (blue circles) to manipulate the movement of groups of muscles that usually work in concert. The middle image shows the bone-driven, muscle system of anatomically accurate muscles. These bones influence the movement of the mesh to display the intended expression as shown in the image on the right.

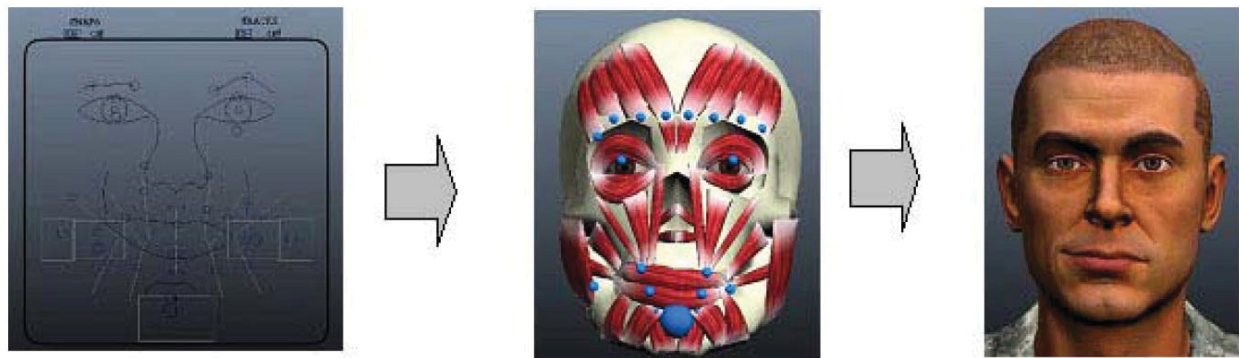


Figure 2. Muscle-based facial rig

The FACS system can also be used to validate that the avatar's expression conveys the same emotion as a human. First, we generate the FACS score for a human subject or reference photo which demonstrates the target expression. Independently, we generate a FACS score for the avatar's expression and then compare the results.

Another element essential to facial expressions is the inclusion of believable lip-synching. Traditional methods of lip-synching are often expensive, time consuming, and may still produce results that are distracting for the observer. We have developed a system to automatically generate a correctly timed lip-synch animation for an audio file of recorded speech based on the corresponding phoneme sequence. Phonemes are the smallest unit of unique speech sounds. Each phoneme has a visual counterpart or mouth shape called a viseme. Using the same process for creating a facial expression, we have created a validated library of visemes. The resulting animation properly simulates human behavior and can be blended with the underlying facial expressions.

6.0 DYNAMIC AND ENGAGING CONTENT

CCAT's instructional content takes the form of scenarios and simulations that embody cultural dilemmas that have been reported by Warfighters and other stakeholders with experience in conducting such missions as

Humanitarian Aid / Disaster Relief (HA/DR), Counter-Insurgency (COIN), and Provincial Transition Teams (PTTs). Virtual mentors from each of the services, as shown in Figure 3, introduce compelling stories which embed these dilemmas.



Figure 3. CCAT Virtual Mentors

To meet the requirement for dynamic and engaging content, CCAT uses Vcom3D's Plug-and-Play Cultural Avatar (PnPCA) system. As shown in Figure 4, PnPCA integrates the physically expressive virtual human role players with a cognitive/affective model of the characters' behavior, as influenced by cultural values. As described in [9], we have used Soar Technology's Cultural Cognitive Architecture (CCA) to provide characters that not only react to learner decisions, but also pursue their own agendas. The CCA combines Hierarchical Value Maps (HVMs), cultural schema, and values-based appraisals to simulate characters whose behaviors are consistent with the culture being modeled.

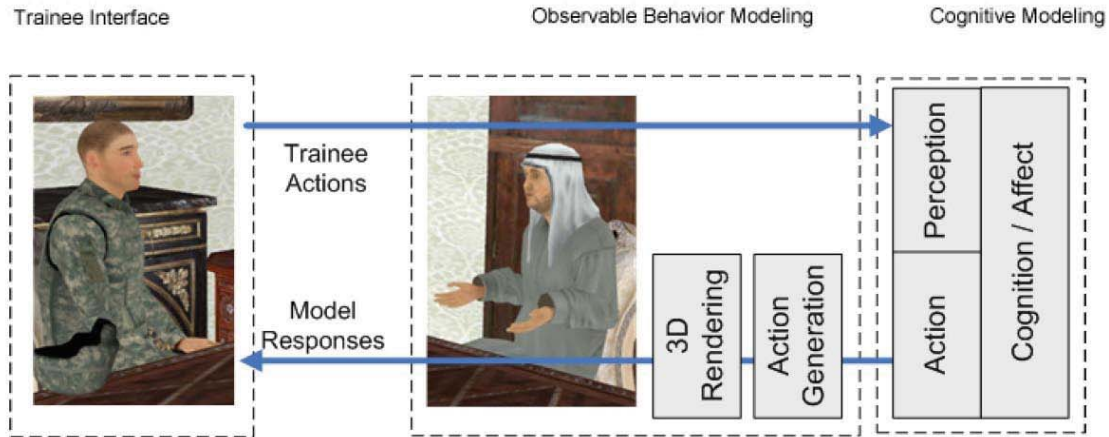


Figure 4. Integration of Physical and Cognitive Model into a Training Environment

7.0 DEVELOPMENT AND DEPLOYMENT

The CCAT process for developing culturally adaptive competency training scenarios begins with a storyboard, which includes the setting, characters, dialog, camera shots, special behavior descriptions, and control flow. Whether the content is to be deployed on the Web, PC, or a mobile platform such as iPhone or Android, the content creation process is nearly the same.

The Content Creation Pipeline, as shown in Figure 5, begins by recording the dialog as performed by native speakers. An animation script author imports each audio file along with the transcript into Vcommunicator Studio which automatically creates the lip-synch animation for the expressive avatar. Vcommunicator Studio is a tool we developed to expedite the character animation process. It allows high-level users, such as subject matter experts or instructional system designers, to directly create the behavior animation sequences. Users work with an intuitive, timeline-based interface to composite body gestures, facial expressions, head and eye tracking, with the lip-synched audio. Studio provides libraries of body gestures which can be filtered by cultural region. The gestures may be layered on top of one another to

combine for example the lower body of one gesture with the upper body of another. New upper body gestures can be created with Gesture Builder, imported into Studio, and composited with existing gestures. The user aligns the resulting behaviors to the dialog to create a completed animation sequence for this character.

The animation is then exported for Autodesk's 3ds Max modeling and animation software where it is automatically applied to a reference character. 3ds Max is also used to create other 3D assets such as the environments, vehicles, props, and any special animations.

As shown in the Game Development Pipeline of Figure 5, the 3D assets are prepared for a game development tool. Once such tool is called Unity and is capable of deploying to a wide range of platforms including Android, iPhone, Windows, Mac, and web browsers. Unity includes an integrated development environment (IDE) where 3D assets are connected and controlled through code and an optimized real-time 3D renderer that uses standard graphics libraries.

Within the IDE, the cultural scenario's game logic is implemented. The game logic will control the flow of the scenario, the character movement, and the camera to

form a cohesive user experience. It will also handle user interaction and trigger the behavior animations for each character.

The final step is to deploy the integrated scenario for the target Operating System.

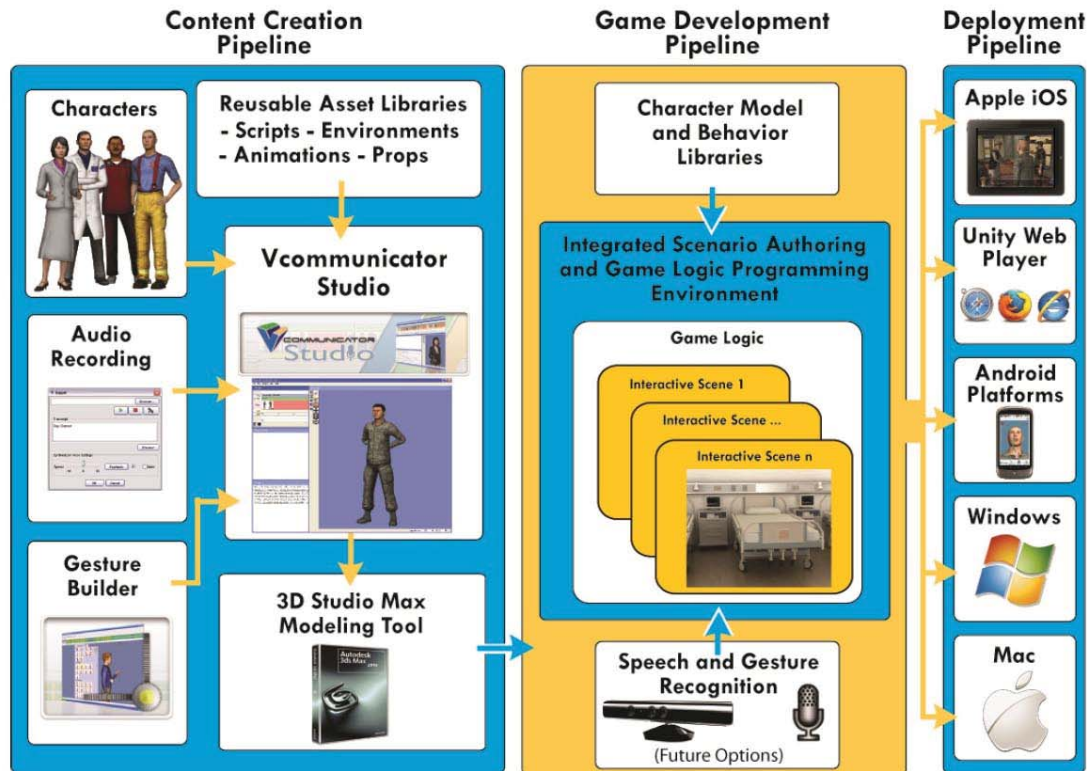


Figure 5. Development and Deployment Pipeline

8.0 EXAMPLE SCENARIO

Utilizing these expressive avatars, we use the described development process to create interactive scenarios depicting authentic cultural dilemmas. The player/learner assumes a defined role with a specific mission objective. Figure 6 shows an example scenario between the player, who assumes the role of a soldier, an Afghan Army officer, who is serving as his interpreter, and a local Afghan Elder. When the player reaches the cultural dilemma with the other stakeholders in the scenario, they are prompted to choose between multiple courses of action. The other stakeholders must be able to respond realistically to enable the student's task of identifying the

cultural values dimensions involved in this dilemma.



Figure 6. Example Scenario

The player must also discern the particular values orientation of each of the other stakeholders. Figure 7 shows the cultural values dimension of Ascription vs.

Achievement and the assignment of the stakeholders' values orientation. In the style of modern "reality show" story-telling techniques, the stakeholder's reaction may be augmented with an "after the fact" interview shown immediately after the reaction. Finally the player is given the opportunity to reflect upon their decision and even observe the result of alternate choices.



Figure 7. Mapping Cultural Values

After working through a number of scenarios in which they learn to recognize and respect cultural differences, the learner then has the opportunity to practice these skills in a mission simulation. These simulations require the learner to discern the cultural values of the virtual characters and to use this understanding to successfully complete missions by resolving or reconciling differences.

9.0 CONCLUSION

In this paper we have described the design considerations that we have applied to develop the CCAT system for cultural adaptivity training. This system combines authoring tools and a multi-platform delivery system for embedding expressive computer-generated characters within a serious games scenario. The system facilitates several training objectives not possible with other methods of learning including:

- Learning thorough variable mission outcomes
- Recognition of non-verbal cues

- Inference of cultural elements through observation
- Recovery from maladaptive behavior
- Time sensitivity - performance of actions in a timely manner and in an appropriate sequence

10.0 REFERENCES

- [1] Glover, G., Friedman, H., Sims, E., & Silvergate, D. (2011). Developing Culturally Adaptive Competencies: A Dilemmas Based Approach.
- [2] Klein, G.A. (1989). Recognition-primed Decisions. In W.B. Rouse (Ed.), *Advances in man-machine systems research* (Vol. 5, pp. 47-92). Greenwich, CT: JAI Press.
- [3] Zbylut, M., Metcalf, K., et al. (2009). *The Human Dimension of Advising: An Analysis of Interpersonal, Linguistic, Cultural, and Advisory Aspects of the Advisor Role*. US Army Research Institute for the Behavioral and Social Sciences, Technical Report 1248. June 2009.
- [4] Sims, E., and Pike, W. (2004). Reusable, Lifelike Virtual Humans for Mentoring and Role-Playing. In 2004 *Proceedings of Interservice/Industry Training Simulation & Education*.
- [5] Sims, E. (2005). Simulating Believable, Context-aware and Culture-specific Human Behaviors. in 2005 *Interservice/Industry Training Simulation & Education Conference*. Orlando, FL.
- [6] Ekman, P. (1971). Universals and cultural differences in facial expressions of emotions. in *Nebraska symposium on motivation*. Lincoln, NB: University of Nebraska Press.
- [7] Matsumoto, D., S. Takeuchi, S. Andayani, N. Kouznetsova, and D. Krupp, The contribution of individualism-collectivism to cross-national differences in display rules. *Asian Journal of Social Psychology*, 1998. 1: p. 147-165.

[8] Ekman, P. and Friesen, W. (1976). Measuring Facial Movement. Journal of Nonverbal Behavior, 1976 - Springer

[9] Sims, E., & Taylor, G. (2009). Modeling believable virtual humans for interpersonal communication. Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), 2009, Orlando, FL.

11.0 ACKNOWLEDGEMENTS

The CCAT training concept described in this paper was developed as a collaboration between Vcom3D and our co-author consultants, Drs. Jerry Glover and Harris Friedman.

This CCAT authoring and delivery system is an extension of the Plug-and-Play Cultural Avatar system, based in part on work performed under US Government contract W91WAW-08-C-0052, monitored by the Army Research Institute and US Government contract W91CRB-09-C-0004, administered by the Combating Terrorism Technical Support Office (CTTSO). Opinions expressed in this paper are those of the authors and do not necessarily represent the views of these sponsoring organizations.

3.3 Using Agent Based Modeling (ABM) to Develop Cultural Interaction Simulations

Using Agent Based Modeling (ABM) to Develop Cultural Interaction Simulations

Nick Drucker
MYMIC, LLC.

nick.drucker@mymic.net

Phillip N. Jones
MYMIC, LLC.

phillip.jones@mymic.net

Abstract: Today, most cultural training is based on or built around “cultural engagements” or discrete interactions between the individual learner and one or more cultural “others”. Often, success in the engagement is the end or the objective. In reality, these interactions usually involve secondary and tertiary effects with potentially wide ranging consequences. The concern is that learning culture within a strict engagement context might lead to “checklist” cultural thinking that will not empower learners to understand the full consequence of their actions. We propose the use of agent based modeling (ABM) to collect, store, and, simulating the effects of social networks, promulgate engagement effects over time, distance, and consequence. The ABM development allows for rapid modification to re-create any number of population types, extending the applicability of the model to any requirement for social modeling.

1. INTRODUCTION

Most military thinkers agree that cultural competency is and will continue to be an essential skill on the part of our service men and women, a skill that has been lacking. The United States and its partners have worked hard to create the training capabilities to fill this gap. For many reasons, most of these current capabilities use training conditions defined by a “cultural engagement” or a discrete dealing between the individual learner and the cultural “other” or others. They are discrete because all trainee attention is focused on a determinate outcome tied to particular training objectives. In reality, however, cross-cultural interactions in the current operational environment involve secondary and tertiary effects with potentially wide ranging consequences. The concern is that learning culture within a strict engagement context leads to “checklist” cultural thinking that will not empower learners to understand the full

consequence of their actions nor tie their cross-cultural success to larger, more strategic aims.

The authors of this paper propose an evolved approach to culture training, one that is “campaign” based, with the potential to be an individual or collective and unit level training tool, and a method for its execution founded on agent based modeling (ABM). Such an approach would include multiple cross-cultural engagements and extend their consequences across distance and time. Such an approach would develop a broader, more strategic perspective on cross-cultural conditions while learning to achieve mission success in cross-cultural conditions.

As the authors have experimented with this new approach within serious games, the paper will use that modality for discussion, though the approach is not limited to game-based solutions.

2. ENGAGEMENT VS. CAMPAIGN

Today, military operational success must be achieved “amongst the people” [10] and cultural competency is essential to achieving that success [5]. One way of delivering culture training is through game-based training systems. Such systems provide an efficient and engaging learning environment and outcomes. Game-based training systems promote learning retention through engagement and require minimal overhead compared with systems dependent on instructor time, materials, and brick and mortar facilities.

There are already several efforts leveraging games for culture training. Some include BiLAT™, Tactical Iraqi™, and Virtual Cultural Awareness Trainer (VCAT™). These efforts share characteristics, a primary being that they are focused on or built around a “cultural engagement” or discrete interactions between the individual learner and one or more cultural “others”. Further, because simulated engagements tend to serve as merely a platform to demonstrate and teach aspects of cultural knowledge, they tend to be highly discrete with extremely limited consequences. In reality, however, cultural interactions are not ends by themselves but rather means to achieve military or other objectives. Additionally, these interactions cause secondary and tertiary effects with potentially wide ranging consequences. Therefore, individuals and groups who are effective in achieving goals under cross-cultural conditions are the focus of true operational objectives. The concern is that learning culture within a strict engagement context might lead to “checklist” cultural thinking that will not empower learners to understand the full consequence of their actions nor foster the initiative sought by the military [13].

The Department of Defense defines engagement as “A tactical conflict, usually between opposing lower echelon maneuver forces” [7]. Within the taxonomy of operations are, in order, engagements, battles, campaigns, and wars. The DoD defines a

campaign as “A series of related major operations aimed at achieving strategic and operational objectives within a given time and space” [7]. Because of the nature of current operations, the campaign is a better context for envisaging cross-cultural performance than the engagement. The Marine Corps Operating Concepts for a Changing Security Environment (2007) [15] identifies that our likely military opponents, unable to achieve decisive results against the United States, will likely leverage duration as a means to achieve their aims. This will result in “prolonged operations.” Further, the Operating Concept discusses the predominance of insurgencies in the past half century and into the future. These are wars among the people and for the support of the people. These same concepts are seen as a basis for current operational theater requirements, as the Commander of ISAF spells out in his Counterinsurgency Guidance (2010) [10]: secure and serve the population, live among the people, hold what is secured, foster lasting solutions, be a good guest, consult and build relationships, fight the information war, manage expectations, and maintain continuity through unit transitions. These are requirements to engage in a prolonged interaction with and for the people where cross-culture is a condition for long-term success.

It is well understood that training is best when it matches actual application: the “train as you fight” paradigm. Learning culture and language in a manner that enables one to apply what has been learned to a broad set of possible situations requires a new approach. Attempting to learn culture through discrete engagements will tend to limit training to heuristics that will be difficult to apply in the complexity of actual operations. Furthermore, it will miss the real target—achieving mission success under cross-cultural conditions.

The documentary Restrepo [12] illustrates the risks of training cultural interaction strictly within the “engagement” context. In the documentary, the American company commander conducts what appears to be a well-coordinated Shura (gathering of village

elder's in Afghanistan). However, in the Shura, the commander makes a point of negatively comparing his predecessor to himself. As the senior representative on the ground this could be perceived as disparaging the International Security Assistance Force (ISAF) or America in general. Had the company commander been trained to see cultural interaction as a campaign versus a series of discrete engagements, he would have better understood the linkages between his actions and those of his predecessor and possibly have had a more effective strategy.

Current virtual environment-based culture trainers tend to employ a single tier of consequence when interacting with cultural-other characters. This is not representative of the real world; in reality, consequences of one's actions will propagate along social networks throughout and beyond communities, having effects across a depth of time, distance, and consequence. These actions will affect future interactions with individuals a trainee has not yet met as well as affecting the interactions of "flanking" individuals and organizations.

What would be required to create a campaign-based approach to cross-cultural training? The authors propose changes to how we model individuals, the cultural others, and new methods to modeling the collective community.

3. MODELING INDIVIDUALS

Humans are inherently complex and difficult to accurately model. To deal with this complexity, or possibly to side-step it, many traditional systems have used intricate branching logic. This solution, unfortunately, leads to stilted models, unable to reflect the dynamics of the human condition.

At the individual level, a thorough framework must include the three domains of human existence: physical, cognitive, and emotional. Physical and cognitive modeling is relatively simple and exists in most first-person games. Emotional struc-

tures, such as the Global Structure of Emotion Types [9], have been developed and could be leveraged to develop computer frameworks that will allow response by AI-driven NPCs.

Sociology and philosophy are also underused disciplines for providing a framework that allows modeling of cross-cultural dynamics. Frame analysis, as introduced largely by Erving Goffman (1974) [6] has provided a basis for decoding how people understand and react to situations. The Social Theory of Practices is another approach for appreciating the natural behaviors of individuals [2] [3] [14] [11]. These theories, taken together, postulate the presence of context or environment to which individuals bring differing frames of reference and personal and cultural practices. The theories illuminate why an office worker acts differently in an office setting than in a social setting or why a local in a foreign community would respond differently than an American based upon a frame of self-identity or honor. These theories provide a means of modeling a much richer and more complete cultural-other, one who responds more complexly and realistically to changed context or predominant frame.

The next logical step is then to tie these more complete individual models to an accurate model of the collective community.

4. MODELING THE COLLECTIVE

The human community is comprised of complex systems. These include networks of people whose interactions result in observable and noteworthy behaviors. People live their lives based on rules (norms), interacting with one-another. Combining groups of people into observable groupings (i.e. a neighborhood, towns, provinces, etc.) provides insight into people composed systems-of-systems. Scaling the systems up creates larger cross sections of society that in turn provides insight on how that society functions. Observing massively

complex human systems requires a set of tools suited for recreating individual behavior based on a known set of rules. One tool is Agent Based Modeling (ABM). ABM is a method whereby experimenters construct complex systems consisting of numerous individual entities called agents [4]. These agents follow a set of rules and interact with each other as they mimic the inner workings of complex systems [8]. Using ABM, researchers re-create extremely complex social systems for study.

Researchers traditionally use ABM as a tool to examine and analyze complex societies. However, the authors propose using ABM as a societal model to capture, store, and propagate the results of cross-cultural engagements, or as the bond that links these engagements across space and time to enable a cross-cultural campaign.

ABM is traditionally used to observe emergent behaviors of groups of individuals. Emergent behaviors are the aggregate of the behavior of all the agents in the ABM system; behaviors which are unforeseen because we cannot normally observe them or notice them in the real world (i.e. one person dropping a rock in a lake will result in a small ripple, but 500 people performing the same act would result in a significant sized wave). Researchers model a targeted population to include means for individual agents to interact, create an initial state, and then run the system, observing the emergent behaviors of the modeled agents. Researchers can subsequently modify agent states and run additional tests, building a set of relevant results to gain insight on the targeted population.

In the approach offered by the authors, the targeted, cultural-other community is modeled within the ABM with several designated agents representing specific characters from the game environment. These characters are programmed to possess the same initial, attitudinal states as their game environment characters. The results of the trainee's engagement will modify the characters' attitudinal states and the system will relay this change to the ABM counterparts. The system cycles, allowing the initial

changes to propagate across the ABM population, changing the states of other individual agents. The system then transfers these subsequent changes back to the virtual environment in terms of changes to the attitudes, beliefs, and behaviors of the characters that trainees will subsequently engage. In a training application, these behaviors become the source of new information to the trainee.

These changes will modify future interactions at the local community level, thus creating a dynamic, consequence included environment for trainees; a trainee's actions will have future ramifications for themselves but also, because ABM social networks will span separate local communities, those actions will have ramifications for neighboring units. This unique expansion of impact over time, space, and consequence will create the conditions for a collective, campaign-based approach to culture training versus an engagement-based approach.

The ABM will also allow a management system that will capture trainee actions and resultant system reactions in a manner that will enable the learner to view the impacts of each interaction as it propagates through the broader social network. This will enable runtime or after action review of the impacts of actions serving as a feedback mechanism.

5. AGENT-BASED MODELING REQUIREMENTS

In order to create a system where cultural-other agents in the ABM can pass changes to the rest of the population, it is important to model accurate social networks within the ABM. Social networks are complex groups of individuals socially connected through some form of communication [1]. For example, a group of people who attend a church, temple, or mosque together are independent of one another, yet they routinely gather to worship together, and in this way share information. This social network is, in turn, part of a larger

network of social networks. ABM is highly suited to modeling these networks to duplicate information transfer, or to accurately replicate the mechanisms of information transfer.

ABM, in this context, is proposed as a tool with which a training system can set and modify interactional / relationship parameters in real time to effect both immediate and extended dynamics between the learner and the virtual cultural-other community. In addition, game-based systems, which require the presentation of learning experiences tailored to the trainee, are more efficient if those experiences are tailored in real or near-real time through assessment of the trainee's responses.

6. CONCLUSION

Cultural competency for today's soldiers consists of prolonged interactions

7. References

- [1]. Barabási, A., (2002), *Linked: The New Science of Social Networks*, Cambridge, MA, Perseus Publishing.
- [2]. Bourdieu, P. (1972), *Outline of a Theory of Practice*. Translated by Richard Nice, New York, NY, Cambridge University Press
- [3]. Bourdieu, P. (1980). *The Logic of Practice*, Translated by Richard Nice, Stanford, CA., Stanford University Press
- [4]. Epstein, J., (2006), *Generative Social Science Studies in Agent-Based Computational Modeling*, Princeton, New Jersey, Princeton University Press.
- [5]. Gates, R., (2011), *Speech Delivered to the United States Military Academy, 25 Feb 2011*, accessed from <http://www.defense.gov/speeches/speech.aspx?speechid=1539>, 8 Apr 2011.
- [6]. Goffman, E., (1974). *Frame Analysis: An Essay on the Organization of Experience*, Boston, MA, Northeastern University Press.
- [7]. Joint Chiefs of Staff 1-02, (2010), *Joint Publication 1-02, Department of Defense Dictionary of Military and Associated Terms*, Joint Chiefs of Staff, Washington, D.C.
- [8]. Miller, J. H., & E. Page, S. E., (2007), *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton, New Jersey, Princeton University Press.
- [9]. Ortony, A., Clore, G. L., & Collins, A., (1988), *The Cognitive Structure of Emotions*, New York, Cambridge University Press.
- [10]. Patraeus, D. (2010), *COMISAF's Counterinsurgency Guidance*, HQ, Interna-

tional Security Assistance Force, Kabul, Afghanistan

- [11]. Schatzki, T. (2002). *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*, University Park, PA, Pennsylvania State University Press.
- [12]. Sebastian J., and Hetherington, T. (Co-Filmmakers). (2010). *Restrepo: One Platoon, One Valley, One Year* [Documentary], Location: Outpost Films and National Geographic Channel
- [13]. Shamir, Eitan (2011), *Transforming Command: The Pursuit of Mission Command in the U.S., British, and Israeli Armies*, Stanford Security Studies, Stanford, CA
- [14]. Turner, S. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presumptions*. University of Chicago Press; Chicago, IL.
- [15]. USMC (2007) *Marine Corps Operating Concepts for a Changing Security Environment*, Commanding General, Marine Corps Combat Development Command Deputy Commandant for Combat Development and Integration, June 2007

3.4 Cognitive Network Modeling as a Basis for Characterizing Human Communication Dynamics and Belief Contagion in Technology Adoption

Cognitive Network Modeling as a Basis for Characterizing Human Communication Dynamics and Belief Contagion in Technology Adoption

Clayton Hutto

Georgia Tech Research Institute
Clayton.Hutto@GTRI.GaTech.edu

Erica Briscoe

Georgia Tech Research Institute
Erica.Briscoe@GTRI.GaTech.edu

Ethan Trehwitt

Georgia Tech Research Institute
Ethan.Trehwitt@GTRI.GaTech.edu

Abstract. Societal level macro models of social behavior do not sufficiently capture nuances needed to adequately represent the dynamics of person-to-person interactions. Likewise, individual agent level micro models have limited scalability – even minute parameter changes can drastically affect a model's response characteristics. This work presents an approach that uses agent-based modeling to represent detailed intra- and inter-personal interactions, as well as a system dynamics model to integrate societal-level influences via reciprocating functions. A Cognitive Network Model (CNM) is proposed as a method of quantitatively characterizing cognitive mechanisms at the intra-individual level. To capture the rich dynamics of interpersonal communication for the propagation of beliefs and attitudes, a Socio-Cognitive Network Model (SCNM) is presented. The SCNM uses socio-cognitive tie strength to regulate how agents influence—and are influenced by—one another's beliefs during social interactions. We then present experimental results which support the use of this network analytical approach, and we discuss its applicability towards characterizing and understanding human information processing.

1.0 INTRODUCTION

The idea of a social network has been attracting interest from researchers in the social and behavioral sciences since the turn of the twentieth century when social theorist George Simmel discussed the implications of individuals' affiliations with groups of others (which he called social circles) [1]. The notion of investigating relationships among entities, and the capability for describing the patterns and regularities of these relationships with precise formal definitions, is indeed appealing. The presence of regularities and patterns in relationships is referred to as structure, and quantities that measure structure are called structural variables [2]. Based on the mathematical concepts associated with graph theory, the network perspective characterizes relationships in terms of nodes and ties. Nodes are the individual actors (e.g., people or organizations) which are tied (or linked) by one or more specific types of interdependent connections such as friendship, kinship,

common interest, financial exchange, or information transfer, for example. Social network analysis is more than just a methodological approach, a convenient vocabulary, or intuitive metaphor for discussing social and behavioral relationships. The network analysis perspective presents a theoretical alternative to the assumption of independent social actors – an assumption that is prevalent in previous sociological and psychological research. The network perspective offers a common framework for testing theories about structured social relationships, and provides a means to precisely characterize important social concepts with explicit formal definitions (see the introductory chapters of [2] for a brief overview).

Unfortunately, as its name implies, the majority of social network research has focused almost exclusively on social structures. Furthermore, much of this research conveys a static, rather than dynamic, social structure. These static views of social structure tend to marginalize the role

of social and cultural transmissions described by Tomasello (i.e., the passing of knowledge via multi-generational history, as well as personal ontogeny, c.f., [3]) when studying human social-cognition interactions. Static network views also poorly capture the situated nature of cognition as put forth by Lave (i.e., dynamically evolving cognition that both influences – and is influenced by – the contextual situation, c.f., [4], [5]). Finally, many of these views do not specify how changes to society occur relative to the individuals' cognition and behavior.

The remainder of this paper is organized as follows: first, we present the Cognitive Network Model (CNM) as a framework for quantitatively characterizing individuals' belief systems as a network of interrelated proposition nodes, with each node having specified quantitative parameters. Next, we present the Socio-Cognitive Network Model (SCNM), which is intended to capture the effects of interpersonal communication and influence on individuals' Belief Networks (BNs) during person-to-person interactions. We then present some preliminary results of computational modeling and simulation studies of the CNM and SCNM, using *technology adoption* as a domain of demonstration. Finally, we discuss how this multilevel modeling approach helps to bridge the divide between micro- and macro- models of human behavior (c.f., [6], [7]), as well as the social and cognitive perspectives of human communication dynamics.

2.0 THE COGNITIVE NETWORK MODEL

The Cognitive Network Model (CNM) is a computational modeling approach to characterize specific cognitive mechanisms associated with human social-cultural information processing at the individual agent level. Similar to Carley's conception of constructivism [8], the fundamental tenet of CNM is the application of network precepts and analysis techniques as a basis for characterizing and understanding human information processing resulting from the diffusion of information. In essence, CNM represents human information processing in

terms of the emergent interactions of a set of parameters associated with beliefs. Thus, beliefs are the foundational element of cognitive network modeling.

2.1 Beliefs: Propositions and Parameters

Beliefs are often represented as propositions and can be thought of as subjective probability estimates of an object having a particular attribute [9], [10]. In the network perspective, belief propositions are represented as network nodes, with links representing relationships such as correlations between the propositions. In the CNM approach, a belief proposition is described as a pairing of cognitive concepts. For example, the proposition "raisins are healthy" pairs the concept of "raisin" with "healthy". The human agent assigns a subjective value of perceived truth to the proposition in order to reflect that individual's level of agreement with the proposition. To represent beliefs, and to provide insight into how to predict the effect of new information, CNM represents beliefs through three main quantitative parameters: veracity, epsilon, and defense (see Figure 1).

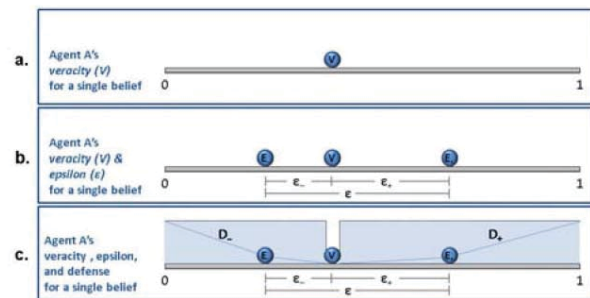


Figure 1: a. **Veracity (V)** is the quantitative parameter in CNM that captures an agent's level of acceptance or rejection in the "degree of truth" of a belief proposition; b. **Epsilon (ϵ)** captures the interval limits on either side of current veracity for which new or alternative values of veracity may be most readily accepted; c. **Defense (D)** reflects the degree of resistance to adoption of alternative veracity values for a given belief proposition.

Veracity (V). Veracity indicates an agent's level of acceptance or rejection in the "degree of truth" of a belief proposition (Figure 1a). Quantitatively, veracity is represented as

a value between 0 and 1, where 0 indicates the agent believes the proposition is not true at all (proposition is completely rejected), and 1 indicates complete belief in the truth of the proposition (proposition is completely accepted). This quantitative characterization of the veracity of the belief overlaps quite well with existing characterizations of beliefs represented as propositions with a subjective probability of being correct [9], [10].

Epsilon (ϵ). Epsilon captures the interval limits on either side of current veracity for which new values of veracity may be most readily adopted (Figure 1b). For a given belief proposition, new values of veracity might be based on information received from an external source such as another agent, contextual evidence, or personal experience. Epsilon is derived from literature based on bounded confidence models of belief formation (c.f., [11-13]), and is a quantitative construct that captures the qualitative concept of certainty (or conversely, uncertainty) for a selected belief proposition. Epsilon also accounts for the qualitative construct of *social proof* (c.f., [14]), the phenomena whereby individuals rely more heavily on interpersonal social influence in conditions of uncertainty. Quantitatively, $\epsilon+$ is the magnitude of the distance between V and the upper boundary limit; $\epsilon-$ is the magnitude of the distance between V and the lower boundary limit. Epsilon may not be uniform in each direction; changes in veracity may be more readily adopted for one direction versus the other.

Defense (D). Defense can be thought of as the degree of resistance to adoption of alternative beliefs (Figure 1c). Quantitatively, defense is a value between 0 and 1, where 0 indicates a weak strength of defense value (i.e., weak resistance to adopting new veracity values), and 1 indicates a strong defense (high resistance, or intolerance). If an agent's defense of the belief is low, differing or conflicting information is less likely to be rejected outright without first modifying one or more belief parameters. Thus, the defense value affects the degree to which an existing belief is amenable to change via (a) adjusting

epsilon, (b) adjusting the strength of veracity, (c) adjusting the strength of the defense, or (d) a combination of belief parameter adjustments. The defense parameter accounts for the qualitative construct of cognitive consistency (or conversely, cognitive dissonance – see [15]) within the belief network, and is the mechanism by which the CNM is able to model both rational and irrational cognitive behavior.

Cognitive network modeling is then a matter of (a) selecting the beliefs of interest to represent a structured belief system for the chosen domain (termed a Belief Network, BN, described in the next section), (b) defining the relevant agent-specific belief parameter values (V , ϵ , and D) for each belief represented in the BN, and (c) describing the governing functions for dependencies between beliefs as well as the role of other intra-personal influences (e.g., age, gender).

2.2 The Belief Network

The Belief Network (BN) is conceptualized according to principles familiar to areas of graph theory and network science: vertices represent individual belief propositions, and edges represent connections/relationships between those beliefs; they are usually weighted and may be either directed or undirected (see [16] for an introduction to graph theory and networks). The BN is the collection of all of the relevant belief propositions for the selected domain (with their associated parameters) held by an individual agent. For example, the BN used as a domain of demonstration in the current effort (see Figure 2) was derived from research literature on the Unified Theory of Acceptance and Use of Technology (UTAUT)[17] that integrates eight of the most prominent theoretical models of technology adoption. Figure 2 depicts the BN containing belief propositions related to performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), voluntariness of use (VoU), behavioral intentions (BI), usage behavior (UB) based on usage experience, and perceptions of trust

and power status of another agent as an information source:

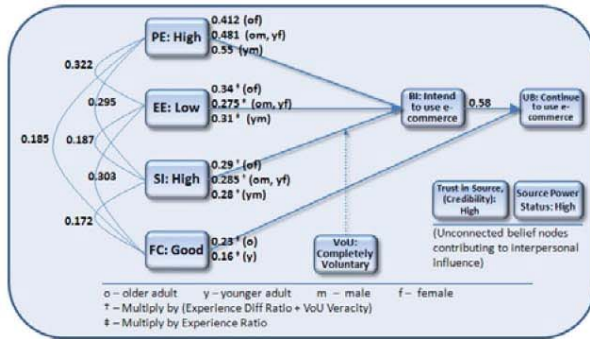


Figure 2: Conceptual rendering of the agent belief network for adoption of technology (e.g., e-commerce), as derived from [17].

From this framework, it follows that CNM is able to capture the concept of culture by adjusting one or more of the following in the BN: (1) the relative strengths of individual belief parameters for belief nodes determined to be culturally germane, (2) the weighted magnitude of the BN edges, and/or (3) the relationships defining the intra-network and external influence governing functions (more on these governing functions in the SCNM discussion, below). Likewise, the BN belief parameters and edge weights can also be initialized according to expected “roles” or “norms”. For example, in a model of technology adoption, we might categorize population segments according to Rogers’ description of the diffusion of innovations [18], whereby individuals’ BN are stochastically drawn from population “types” such as innovators, early adopters, early majority, late majority, laggards, and luddites. It is recognized that an individual human agent’s BN (representing cognition) interacts with other, non-cognitive personal determinants such as personality, transient affective states (moods), biological factors (e.g., age, gender), and personal history or experiences (e.g., education, work). Although not detailed herein, they are considered an integral part of the CNM, and need to be notionally represented in the agent based model. However, it should be noted that the agent-based computational model used for the

research described in this paper only partially accounts for these individual differences.

If beliefs are conceived as subjective valuations of the degree of acceptance or rejection of information (propositions), then the model needs to support changes to targeted belief propositions resulting from interpersonal influence. This may be accomplished either through shifts in strength of veracity of the belief (“raisins actually are not very healthy”), shifts in epsilon surrounding the veracity of the belief (“I’m no longer certain how healthy raisins really are”), or changes in the strength of the defense of the belief (“If I trust someone as a credible source of new information, I may not resist adopting their view that raisins are unhealthy”). Thus, the belief parameters of veracity, epsilon, and defense are the mechanisms by which cognitive network modeling quantitatively characterizes many of the processes associated with human information processing. These processes are discussed in further detail in the context of interpersonal interactions within the Socio-Cognitive Network Model, described next.

3.0 THE SOCIO-COGNITIVE NETWORK MODEL

The Socio-Cognitive Network Model (SCNM) extends the representation of belief propositions into communication patterns that exist between agents. At the SCNM level, network vertices are individual human agents, and network edges are the cognitive and social ties between them. Socio-cognitive network modeling is extremely relevant to the study of human communication dynamics and interpersonal interactions: just as the BN does not function in isolation within an individual agent, neither do agents act in isolation – a number of social and environmental factors interact with the agent to affect the agent’s cognition. For the purposes of this paper, we focus on factors of interpersonal influence as a moderator of belief propagation in social networks. Specifically, within the SCNM, interpersonal influence is captured by a variable called the Socio-Cognitive Influence Power (SCIP). The SCIP is a critical index of

an agent-node within the SCNM, as it indicates the influence power of an adjacent agent j over the current agent i . As an extension of Granovetter's notion of *tie strength* [19], the SCIP index is determined by considering the relative strength of ties along four dimensions (see Table 1):

Table 1: Four Dimensions of Socio-Cognitive Influence Power (SCIP) index.

1. **Structural dimension:** relative patterns of vertex and edge relationships with regards to network structure (c.f., [20-26])
2. **Cognitive dimension:** relative agreement of Belief Network parameters between agents
3. **Interaction dimension:** relative interaction intensity (i.e., frequency, type/medium, and quality – c.f., [19], [26-34])
4. **Social dimension:** relative perception of the strength of the social relationship (e.g., trust, credibility, and emotional support [35]; social distance [36] or status similarity [37]; and power status [38]).

For the sake of space, we limit the current discourse to discussions primarily regarding the cognitive dimension. A central premise of

SCNM is that the strength of the cognitive ties (degree of similarity between agents' BNs) will affect the degree to which agents influence—and are influenced by—one another's beliefs during social interactions. By representing the relevant beliefs for a given domain, as well as the governing functions associated with the modification of those beliefs, the proposed multi-level model can monitor the changes in adoption or rejection of belief propositions (both at the individual and at the aggregate levels) to determine the effects of information as it propagates through the social network. For example, during interpersonal interactions, the belief parameters can be dynamically adjusted according to certain interpersonal influence factors. This dynamic adjustment is determined via heuristic algorithms which use condition-based governing functions for interpersonal interactions (see Figure 3).

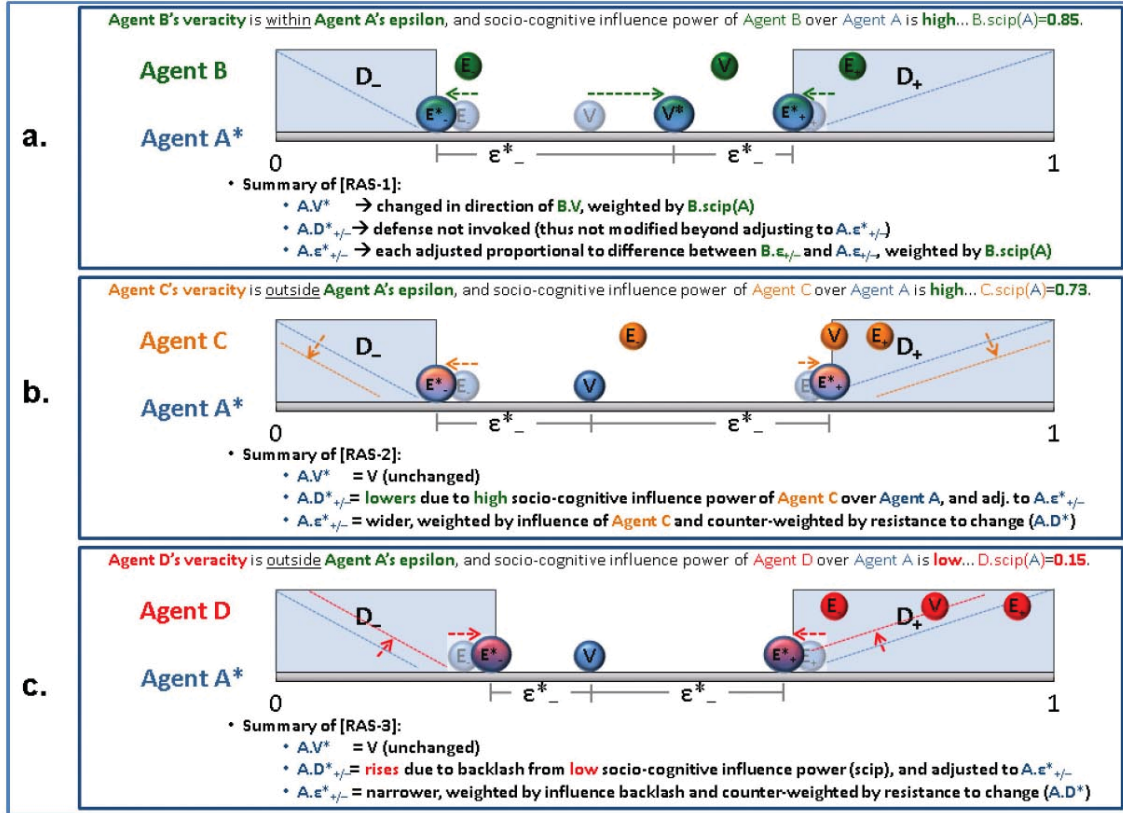


Figure 3: Notional illustrations of example governing functions (reconciliation algorithm sets, or RASs) for the acceptance or rejection of belief propositions during interpersonal interactions.

Figure 3a shows the situation where Agent B is presenting the case for a more extreme veracity value for a selected belief proposition. The proposed veracity is within Agent A's epsilon (i.e., the zone of ready adoption – simulated in this case as the most extreme condition where the defense is not invoked at all [value is 0.0], indicating no resistance to adopting the proposed veracity), and Agent B's SCIP over Agent A is high (0.85). Thus, Agent B influences Agent A to adopt a more extreme veracity value for the belief proposition.

Next, Figure 3b illustrates the situation where Agent B is again attempting to convince Agent A to adopt a more extreme veracity value for a selected belief proposition. This time, the proposed veracity position is just outside Agent A's epsilon, thus the defense parameter is invoked. Agent B does not successfully get Agent A to adopt the more extreme veracity during this interaction. However, because Agent B is a trusted source (SCIP of Agent B over Agent A is high, 0.73), Agent B is successful at making Agent A less confident, less certain about the belief proposition, and lowering Agent A's resistance to change (both of which will make it easier to influence Agent A on subsequent interactions).

Finally, Figure 3c depicts the interesting situation where an attempt by Agent B to influence Agent A actually *backfires* such that the consequence is that the interaction has the opposite effect of what is desired. This occurs in the SCNM as a result of Agent B attempting to convince Agent A to adopt a very extreme veracity position for a selected belief proposition, but not only is the proposed veracity position outside Agent A's epsilon (thus invoking the defense parameter), but Agent B is perceived as being untrustworthy or cognitively dissimilar (SCIP of Agent B over Agent A is low, 0.15) that Agent A actually entrenches in their own belief. Agent A becomes more certain of his own veracity position (i.e., epsilon narrows, perhaps due to Agent A explaining why he thinks his position

is actually superior, or describing shortfalls or other reasons for inferiority of Agent B's position). Agent A's defense (resistance to change) also increases. These will make it even more difficult for Agent B to influence Agent A on subsequent interactions.

4.0 PRELIMINARY RESULTS

Using the Unified Theory of Acceptance and Use of Technology (UTAUT) [17] as the basis for developing the relevant BN (see Figure 2), as well as a number of notional governing functions (some of which are described above), we demonstrate the how the CNM and SCNM models collectively account for a number of phenomena noted in the diffusion literature. Figure 4 illustrates the relationships between the CNM parameters of defense and epsilon (aggregated for all beliefs in the BN for each agent) to the adoption rate of a particular technology (i.e., e-commerce) in a computer simulation study. The relationship shows a general trend in the direction that would be expected according to the CNM concepts – those agents with lower defense and wider epsilon are the “easiest” to influence and were thus the earliest adopters. As defense rises, and as the epsilon range becomes narrower, agents become more “difficult” to influence, thus taking longer to adopt.

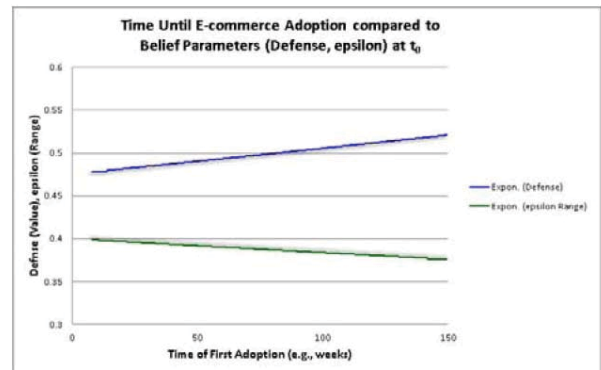


Figure 4: Depicts the relationships between the CNM parameters of Defense and Epsilon (aggregated for all beliefs in the BN for each agent) to the time of adoption of a technology.

Also, Figure 5 illustrates the distinctive processes related to Rogers' first and second

stages of diffusion of innovations [18]. The first stage is the diffusion of awareness of new information (exposure to alternative belief propositions); the second stage reflects the influence processes where by individuals are persuaded to adopt:

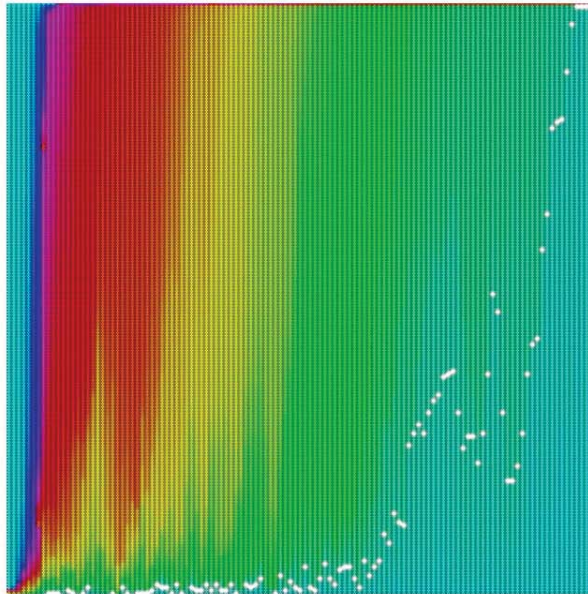


Figure 5: Illustration of the distinctive processes of awareness of new information (alternative belief propositions) and the persuasion process to adopt those alternative views. The horizontal axis represents time while the vertical axis represents individual acceptance. White indicates the transition point at which all subsequent persons have adopted.

5.0 DISCUSSION AND FUTURE WORK

Cognitively motivated theories have the potential for providing a framework that can be used to explain a wide variety of social phenomena. For the most part, however, many models of social behavior have historically been remiss in incorporating cognitive models [8], [39]. This work has presented an approach that uses a multi-level model that exchanges relevant information across representational scales in order to demonstrate the pairing of theoretically grounded models of human cognition with those that account for the influence of interpersonal interactions. The Cognitive Network Model (CNM) and Socio-Cognitive

Network Model (SCNM) is the first and second levels of what is to become a multi-scaled modeling approach that helps to narrow the gap. The proposed models are intended to provide a means of describing and explaining specific processes related to the propagation of information across communication and social networks. A specific issue that this model intends to explain is the occurrence of a backfire effect, where a communication causes a reaction that is an opposite direction of the intended influence (e.g., [40]).

The proposed representation scheme is currently being refined and validated by comparing to 'real-world' data that exists in the domain of technology adoption as well as in human experimentation on belief propagation. These experiments are designed to test a number of hypotheses which emerge from the CNM and SCNM characterization of belief propagation and interpersonal influence. In addition, we intend to incorporate other important metrics into the network models, including other relevant agent-factors, e.g. the emotional intensity of a communication and biological differences in age and gender.

6.0 CONCLUSION

This paper has presented a multi-scale network perspective approach using a framework for representing the beliefs of an individual (represented in an agent-model) as well as the communication and influence on those beliefs by the individual's social interaction. While this work is still in its infancy, we anticipate that planned experiments will provide a validation and refinement of the model that will then allow it to make predictions of human behavior.

7.0 REFERENCES

- [1] G. Simmel, *The Sociology of Georg Simmel*, Compiled and translated by Kurt Wolff. Glencoe, IL: Free Press, 1950.
- [2] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*.

- Cambridge, United Kingdom: Cambridge University Press, 1994.
- [3] M. Tomasello, *The Cultural Origins of Human Cognition*. Harvard University Press, 2001.
 - [4] J. Lave, *Cognition in practice: mind, mathematics, and culture in everyday life*. Cambridge University Press, 1988.
 - [5] J. Lave and E. Wenger, *Situated learning: legitimate peripheral participation*. Cambridge University Press, 1991.
 - [6] J. S. Coleman, *Foundations of social theory*. Belknap Press of Harvard University Press, 1994.
 - [7] C. Goldspink and R. Kay, "Bridging the Micro-Macro Divide: A New Basis for Social Science," *Human Relations*, vol. 57, no. 5, pp. 597-618, 2004.
 - [8] K. Carley, "The Value of Cognitive foundations for Dynamic Social Theory," *Journal of Mathematical Sociology*, vol. 14, pp. 171-208, 1989.
 - [9] I. Ajzen, "Nature and operation of attitudes," *Annual Review of Psychology*, no. 52, pp. 27-58, 2001.
 - [10] M. Fishbein, *Belief, attitude, intention, and behavior: an introduction to theory and research*. Reading, Mass: Addison-Wesley Pub. Co., 1975.
 - [11] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems (ACS)*, vol. 3, no. 1, pp. 87-98, 2000.
 - [12] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence: Models, analysis and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, pp. 1-24, 2002.
 - [13] J. Lorenz, "Continuous Opinion Dynamics under Bounded Confidence: A Survey," *International Journal of Modern Physics*, 2007. [Online]. Available: <http://arxiv.org/abs/0707.1762>.
 - [14] R. B. Cialdini, *Influence: the psychology of persuasion*. Collins, 2007.
 - [15] L. Festinger, *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press, 1957.
 - [16] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, MA: Cambridge University Press, 2010.
 - [17] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425-478, 2003.
 - [18] E. M. Rogers, *Diffusion of Innovations*, 5th ed. New York, NY: Free Press, 2003.
 - [19] M. S. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380, 1973.
 - [20] R. S. Burt, *Structural holes: the social structure of competition*. Harvard University Press, 1995.
 - [21] D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277-293, 1956.
 - [22] N. E. Friedkin, "A test of structural features of granovetter's strength of weak ties theory," *Social Networks*, vol. 2, no. 4, pp. 411-422, 1980.
 - [23] N. E. Friedkin, "Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study," *American Sociological Review*, vol. 58, no. 6, pp. 861-872, 1993.
 - [24] S. A. Golder and S. Yardi, "Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality," in *Proceedings of the Second IEEE International Conference on Social Computing*, Minneapolis, MN, 2010.
 - [25] N. Lin, P. W. Dayton, and P. Greenwald, "Analyzing the Instrumental Use of Relations in the Context of Social Structure," *Sociological Methods Research*, vol. 7, no. 2, pp. 149-166, 1978.
 - [26] J. P. Onnela et al., "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332-7336, 2007.
 - [27] G. K. Zipf, *Human behavior and the principle of least effort: An introduction to human ecology*. Menlo Park, CA: Addison-Wesley Publishing, 1949.
 - [28] R. B. Zajonc, "Attitudinal Effects Of Mere Exposure," *Journal of Personality and*

- Social Psychology*, vol. 9, no. 2, pp. 1-27, 1968.
- [29] L. Festinger, S. Schachter, and K. Back, Eds., *Social Pressures In Informal Groups: A Study of Human Factors in Housing*. Stanford, CA: Stanford University Press, 1950.
 - [30] S. Feld, "The Focused Organization of Social Ties," *The American Journal of Sociology*, vol. 86, no. 5, pp. 1015-1035, 1981.
 - [31] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the 27th international conference on Human factors in computing systems*, Boston, MA, USA, 2009, pp. 211-220.
 - [32] P. V. Marsden and K. E. Campbell, "Measuring tie strength," *Social Forces*, vol. 63, no. 2, pp. 482-501, 1984.
 - [33] K. M. Matthews, M. C. White, R. G. Long, B. Soper, and C. W. Von Bergen, "Association of Indicators and Predictors of Tie Strength," *Psychological Reports*, vol. 83, no. 2, pp. 1459-1469, 1998.
 - [34] C. Haythornthwaite, "Strong, Weak, and Latent Ties and the Impact of New Media," *The Information Society*, vol. 18, pp. 385-402, 2002.
 - [35] A. Petroczi, T. Nepusz, and F. Bazso, "Measuring Tie Strength in Virtual Social Networks," *Connections*, vol. 27, no. 2, pp. 39-52, 2007.
 - [36] N. Lin, W. M. Ensel, and J. C. Vaughn, "Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment," *American Sociological Review*, vol. 46, pp. 393-405, 1981.
 - [37] P. F. Lazarsfeld and R. K. Merton, "Friendship as a social process: a substantive and methodological analysis.," in *Freedom and Control in Modern Society*, M. Berger, Ed. New York, NY: Van Nostrand, 1954, pp. 18-66.
 - [38] J. R. P. French and B. Raven, "The bases of social power," in *Group dynamics*, New York, NY: Harper& Row, 1959.
 - [39] A. V. Cicourel, *Cognitive Sociology*. New York, NY: The Free Press, Macmillan Publishing Co., 1974.
 - [40] Z. L. Tormala and R. E. Petty, "What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty," *Journal of Personality and Social Psychology*, vol. 83, no. 6, pp. 1298-1313, 2002.

3.5 A Systematic Approach for Real-Time Operator Functional State Assessment

A Systematic Approach for Real-Time Operator Functional State Assessment

Guangfan Zhang, Wei Wang, Aaron Pepe, and Roger Xu
Intelligent Automation, Inc.
{gzhang, wwang, apepe, hgxu}@i-a-i.com

Tom Schnell, Nick Anderson, Dean Heitkamp
University of Iowa
thomas-schnell@uiowa.edu; nick@andersonaviation.net; dean-heitkamp@uiowa.edu

Jiang Li, Feng Li, and Frederick McKenzie
ili@odu.edu; feng.odu@gmail.com; rdmckenz@odu.edu

Abstract. A task overload condition often leads to high stress for an operator, causing performance degradation and possibly disastrous consequences. Just as dangerous, with automated flight systems, an operator may experience a task underload condition (during the en-route flight phase, for example), becoming easily bored and finding it difficult to maintain sustained attention. When an unexpected event occurs, either internal or external to the automated system, the disengaged operator may neglect, misunderstand, or respond slowly/inappropriately to the situation. In this paper, we discuss an approach for Operator Functional State (OFS) monitoring in a typical aviation environment. A systematic ground truth finding procedure has been designed based on subjective evaluations, performance measures, and strong physiological indicators. The derived OFS ground truth is continuous in time compared to a very sparse estimation of OFS based on an expert review or subjective evaluations. It can capture the variations of OFS during a mission to better guide through the training process of the OFS assessment model. Furthermore, an OFS assessment model framework based on advanced machine learning techniques was designed and the systematic approach was then verified and validated with experimental data collected in a high fidelity Boeing 737 simulator. Preliminary results show highly accurate engagement/disengagement detection making it suitable for real-time applications to assess pilot engagement.

NOMENCLATURE

ANOVA: ANALYSIS OF VARIANCE

ATC: Air Traffic Controller

ATP: Airline Transport Pilot

BP: Back Propagation

BPS: Boredom Proneness Scale

CATS: Cognitive Avionics ToolSet

CDU: Control & Display Unit

ECG: Electrocardiogram

EEG: Electroencephalogram

GMM: Gaussian Mixture Model

MEL: Multi-Engine Land

NASA: National Aeronautics and Space Administration

NATO: North Atlantic Treaty Organisation

NN: Neural Network

OFS: Operator Functional State

PSD: Power Spectrum Density

SART: Situational Awareness Rating Technique

SEL: Single-Engine Land

SVM: Support Vector Machine

TLX: Task Load Index

1.0 INTRODUCTION

The primary focus of this research is to provide a real-time Operator Functional State (OFS) assessment mechanism. According to North Atlantic Treaty Organisation (NATO) [1], OFS is defined as the multidimensional pattern of human psychophysiological condition that mediates performance in relation to physiological and psychological costs. In aviation systems, two types of hazardous operator states are likely to lead to human errors [2]: a stress state due to high cognitive workload (we do not consider physical workload in this research) or a complacent/bored state due to extremely low cognitive workload over a

prolonged period of time [3]. Proper assessment of cognitive workload and appropriate workload modulation offer the potential to improve mission effectiveness and aviation safety in both overload and underload conditions [3]- [6]. In commercial flights (especially long-haul flights), pilots often experience periods of high workload during pre-flight preparations, takeoff and landing, as well as longer periods of very low workload as the pilot cruises enroute toward the destination with the aircraft on autopilot. Pilots can easily become disengaged during the enroute phase as they may be less attentive under low workload. When unexpected events occur, especially for fatigued pilots, disengagement could lead to operational errors. Such events can include unexpected changes in weather (turbulence, for example), equipment failure/malfunction (such as hydraulic pump failure) or potential collisions with other aircraft.

During the past few years, we have been developing a systematic approach for OFS assessment [7][8]. In our previous research, we have conducted preliminary studies on identifying the ground truth for OFS, which is required to train an OFS assessment model. Data is needed to train an OFS assessment model based on input signals (subjective/objective measures) and associated operator states, such as engaged/disengaged labels. Many existing studies have utilized psychophysiological measurements to index the level of cognitive demand associated with a task [2][3], fatigue [9][10], engagement [11][12], and other functional state dimensions [1]. However, most of them label OFS based on subjective evaluations and are often very sparse temporarily (only a few labels during the whole experiment). In our research, we have identified several types of information sources that can indirectly infer engagement, improving the ground truth finding procedure described in [7] to take into account the degradation and recovery of an OFS due to factors such as changes in workload. At the same time, we

developed an enhanced committee machine-based model for engagement assessment. The developed techniques have been verified and validated with experimental data collected from a Boeing 737 simulator. Initial results show accurate real-time assessment of pilot engagement state.

The remainder of the paper is organized as follows. Section 2 describes the flight simulation configuration and experiment design. Section 3 presents a mechanism to determine the ground truth for engagement modeling. Section 4 describes an enhanced committee machine based real-time engagement assessment model. Section 5 shows preliminary performance evaluation results. Section 6 concludes the paper.

2.0 ENGAGEMENT ASSESSMENT EXPERIMENT DESIGN

In order to study engagement, we conducted experiments in a fully equipped Boeing 737 simulator [13] involving commercial pilots. In an earlier study conducted by Ellis [13], he described the functionality of the simulator as a fully functional flight deck with full glass cockpit displays, five outside visual projectors, functioning mode control panel with autopilot and autothrottle, and standard Boeing 737 controls (Figure 1).



Figure 1. Boeing 737-800 simulator

2.1 Participants

Several subjects have participated in the pilot engagement study. Pilots had varying levels of experience with different types of aircraft. All had an instrument rating and commercial/private/ATP (Airline Transport Pilot) licenses with experiences in Single-Engine Land (SEL), Multi-Engine Land (MEL), Jet or Turboprop. As an example, Table 1 lists experience levels of four subjects and the type of licenses they hold.

Table 1. Example of participant flight experience and licenses

Subject	SEL	MEL	Jet	Turboprop	Licenses
1	45	0	0	0	Commercial
2	15	0	0	0	Commercial
3	7	3.5	2.5	0	Commercial, ATP
4	22	14	14	1	ATP

2.2 Experimental design

The experiments involved a flight from Seattle Tacoma International Airport to Chicago O'Hare International Airport. The details of the flight have been extracted from an actual American Airlines flight which took place on May 10th, 2010. Details were provided on flightaware.com. The flight path is represented by the blue line in Figure 2.

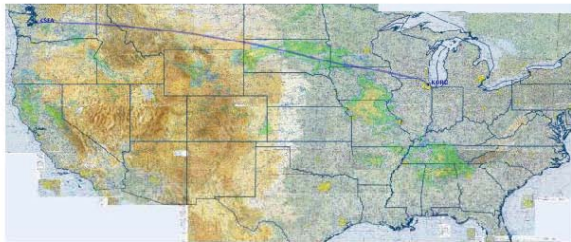


Figure 2. Simulation flight path

In order to study the effects of sleep-loss related fatigue on engagement, all pilots were scheduled to arrive at 5:30pm and were asked to avoid drinking caffeinated beverages such as coffee on the day of the experiment. An orientation video was shown to the subjects before the simulated flight. The video contained a description of the experiment as well as a Control & Display Unit (CDU) programming training section. The video included a description of the sensors and video recording devices used

during the experiment, as well as the responsibilities that the pilots would have during the experiment. The details shared with the subjects did not include information on the probes that were used to measure engagement levels so that the pilots did not anticipate these probes throughout the experiment. During the flight simulation, one of the staff controlled the simulation computer to play pre-recorded audio files mimicking ATC transmissions. An experimenter was in charge of tagging the data to make sure that proper labels were added to the data sheets to identify the phases of the experiment as well as the times when the pilot responded to ATC. At the end of the experiment, the subjects filled out a subjective survey to assess their workload, fatigue and situational awareness during different phases of flight.

The flight simulation included three events inserted into the flight scenario. The events were scheduled to occur at predetermined times to observe and measure how pilots responded. The first event was an ATC call asking the pilot to report when the aircraft was at 29000 feet. This call came while the aircraft was crossing 19000 feet. The aim of this event was to assess whether the pilot would remain engaged at the early stage of initial ascent. The second event was another ATC call that asked the pilot to report their position at 20 miles east of HLN (one of the waypoints). This call came at the early stages of level flight. The goal of this event was to determine whether the pilot would remember to call back ATC at the designated point. The third probe was a failure event. Half of the subjects received a failure signal 1 hour into the flight simulation. The other half received the signal 3 hours into the flight simulation. This approach was preferred because if all runs had both failures, the pilots may have remained in an engaged state throughout the flight after the first failure, with the expectation that such failures may be inserted into the scenario to test his/her performance. The data collected during these events can be compared to establish

the difference in the two engagement states in terms of physiological measures and subjective ratings. The event was selected such that it wouldn't prompt a drastic decision such as an emergency landing but would allow the pilot to solve the problem with onboard capabilities.

2.3 Subjective evaluations

In the experiments to be performed, we included several subjective rating scales that were collected after each experiment, including: Situational Awareness Rating Technique (SART [14]), Bedford workload scale [15], NASA Task Load Index (NASA TLX [15]), Samn-Perilli fatigue scale [16], and Boredom Proneness Scale (BPS) [17].

In order to minimize the effects of intrusive questioning, a post experiment survey was conducted. Each subject was asked about his/her boredom proneness/perceived level of workload, and situation awareness and fatigue during different phases of flight.

2.4 Objective observations

In addition to flight technical data (altitude, speed, etc.), objective data collection was achieved with the use of three types of sensors, including eye tracking cameras, a EEG net, and an EKG sensor. The data was streamed into the Cognitive Avionics Toolset (CATS), which is an analysis tool for operator functional state assessment. A snapshot of CATS is shown in Figure 3.

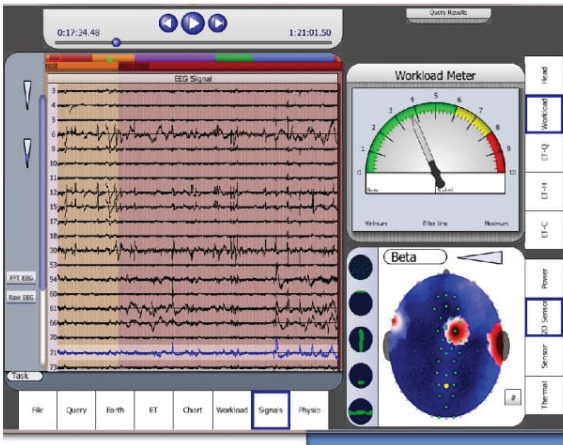


Figure 3. EEG signal view in CATS

In addition, performance data, such as response time to ATC calls or pump failure, were also collected.

3.0 ENGAGEMENT GROUND TRUTH FINDING

Before an engagement assessment model can be deployed, it needs to be trained based on the engagement ground truth and corresponding input information (physiological signals, performance, and others). However, there does not exist a sensor to provide engagement ground truth. In this paper, we created an engagement ground truth assessment model incorporating subjective evaluation, behavioral measures (such as communications with ATC and real time performance data), and sensor measures (such as EEG, eye tracking and EKG data).

The subjective evaluation data was collected after each pilot completed the flight simulation. We divided the whole flight simulation into 11 phases from takeoff to taxi and to gate (plus a special phase when the pilot was handling the pump failure), as shown in Table 2. For each phase, each pilot gave a score for each dimension in SART, Bedford workload scale [15], NASA TLX, and Samn-Perilli fatigue scale. Each pilot also rated his boredom proneness based on the survey.

Table 2. Phases during the flight

Takeoff to 19K
19K to 29K
29K to 37K
Seattle Center
Failure/Seattle Center
Salt Lake Center
Minneapolis Center
Chicago Center
Chicago Center to Call to Descend
Call to Descent to Leveling at 9000
Final Descent to Land
Touchdown and Taxi to Gate

To derive an engagement profile as ground truth for OFS model training, we need to consider different sources of information. Three major steps are followed: baseline construction, degradation/recovery, and refinement based on strong indicators.

1. **Baseline construction.** Engagement baseline is constructed based on possible incentives/motivations. A pilot with strong motivation or in a mission with high incentive usually has a relatively higher engagement level. In the initial study, we set the engagement baseline to a constant highest level for simplicity.
2. **Degradation/recovery.** Engagement status usually changes due to workload/task change and/or occurrence of unexpected events. Expected events include regular ATC calls or corresponding replies. Although those expected events do not have a precise time schedule, their happening would not surprise the pilots. When expected events happen, the operator's engagement level only increases by the minimum amount necessary to accomplish the known task, and decreases to the previous amount quickly thereafter. On the other hand, unexpected events are those that the pilots are not prepared for. In our experiment, the pilots were not aware of the pump failure event in advance. We hypothesize that a pilot has a more rapid engagement recovery when an unexpected event happens, and it can keep the pilot alert for a longer period of time, which indicates a slower degradation speed in engagement level thereafter.

3. **Refinement based on strong indicators.** Strong disengagement/engagement indicators based on measurements, such as eye closure/head drooping (due to fatigue) indicating a disengaged state and short R-R interval (fast heart beat) indicating

an engaged state shall be utilized for engagement refinement. In our initial research, described in this paper, we used pilot's R-R (heart beat) interval as an indicator for engagement level. High R-R interval values imply a relaxed stage in which the pilot's engagement level will degrade, and low R-R interval values indicate an engagement recovery stage.

The degradation/recovery speed of engagement is unique to each individual. This individual difference may be estimated through objective measures, such as an individual's fitness level, or with subjective evaluations, such as the boredom proneness scale. An easily bored operator usually gets distracted faster, and the lower the workload is, the faster the engagement level drops. In summary, the schema is shown in Figure 4.

4.0 ENGAGEMENT ASSESSMENT MODEL

With the engagement level labeled, an engagement assessment model has been trained. The inputs to the engagement assessment model include EEG, eye tracking and EKG data. These signals were first preprocessed (filtering, outlier removal, artifact removal), and the most relevant features were extracted and selected.

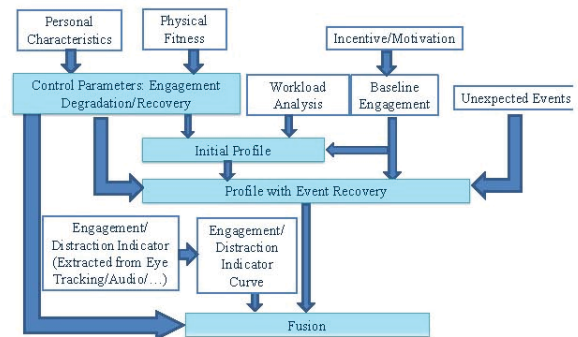


Figure 4. Engagement ground truth finding

Given the input feature sets, a committee machine method has been implemented to relate these features to an OFS. The basic idea of a committee machine [18] is to

aggregate the outputs from several OFS assessment models/committee members specified by users. Different algorithms can function as committee members, for example, Neural Networks (NNs), Gaussian Mixture Models (GMMs), and Support Vector Machines (SVMs). In this research, the committee machine was implemented using a multilayer perceptron trained by the standard Back Propagation (BP) algorithm as the base classification model.

5.0 EXPERIMENTAL RESULTS

The developed techniques were evaluated with the experimental data collected through a Boeing 737 flight simulator. During data collection, a sensor set up problem caused most of the ECG data to become contaminated and this data was not used for the initial evaluation. Next, we describe the results of engagement ground truth finding, data processing, and OFS modeling evaluation results. In [19] (in press), we presented detailed results on sensor data processing, feature analysis, and modeling.

5.1 Ground Truth Finding

The engagement level of operator varies during the entire period of flight and cannot be determined *a priori*. In this paper, we create a benchmark engagement based on commonly accepted assumptions.

Taking off and landing are tasks in which pilots usually are highly engaged since these tasks are relatively more challenging. Also, when pump failures happen and pilots realize the failures, they will be fully engaged due to the emergency condition. On the other hand, during level flight, pilots tend to be at a low engagement level due to flight automation.

Therein, we set a high engagement level (100) at takeoff, landing and pump failure handling phases, and set a low engagement level (30) during level flight periods. A middle engagement level (80) is assigned to climbing and descending tasks. Also, high and medium engagement task periods are followed by a 10 minute extension since it

takes time for the pilot to relax. Figure 5 shows the benchmark engagement during the flight. Although the benchmark engagement is not an accurate prediction, it provides a reference to validate the proposed engagement assessment model.

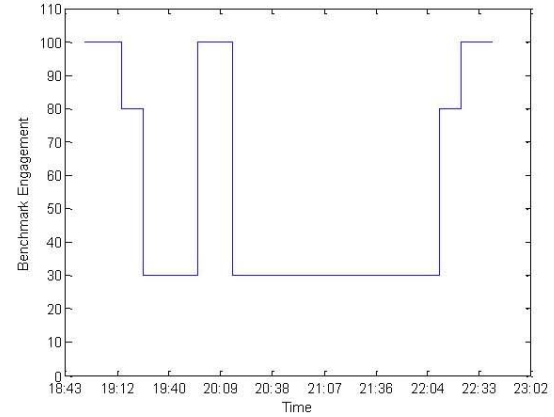


Figure 5. Benchmark Engagement

Figure 6 shows the workload and R-R interval for subject 1 during the flight. High workload values are observed during takeoff and landing phases, as well as during the time spent handling the pump failure. Based on measurements of workload and observation during the experiment, we believe the fluctuations of R-R interval correlate well with and can contribute to the estimation of benchmark engagement

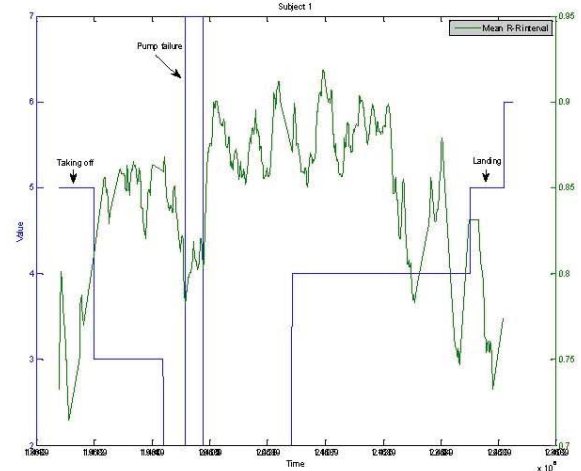


Figure 6. Mean R-R interval (in green) and workload (in blue) during the flight

Based on the workload and heart rate information collected during the experiment, along with expected and unexpected events, our proposed engagement ground truth finding model generates a continuous real-time engagement evaluation, as shown in Figure 7.

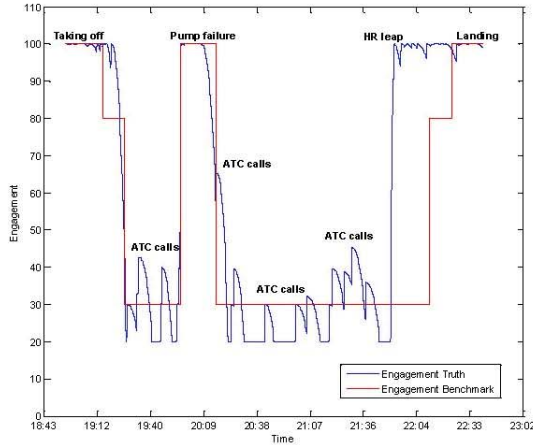


Figure 7. Engagement ground truth

Again, we can see that high engagement level is experienced during takeoff, landing and failure handling phases. Low engagement is experienced during level flight periods. Meanwhile, the ground truth profile reflects the engagement state variation due to expected events, such as ATC calls.

We calculated the Pearson's correlation between the generated engagement curve and benchmark. The correlation coefficient is 0.7894 with a p-value less than $1e-5$, which indicates their correlation with a high confidence level.

5.2 Experiment Data Preparation for OFS Modeling

In this study, two kinds of engagement states and their time durations were first identified by watching videos. A pilot's state during takeoff or while handling a pump failure was considered as 'engaged' (or as state 2) The pilot's state during level flight without any manipulation or if napping was recognized as 'disengaged' (or as state 1). Calculated features can then be labeled

with these states by aligning with the identified time information, as shown in Table 3.

Table 3. Pilots' engagement states

	Duration	State	comments
1	19:08:00 ~ 19:18:00	2 (Engaged)	Taking off
	21:08:00 ~ 21:17:00	1 (Disengaged)	Level flight
2	19:52:00 ~ 20:03:00	2 (Engaged)	Taking off
	21:19:00 ~ 21:29:00	1 (Disengaged)	Level flight
3	19:13:00 ~ 19:23:00	2 (Engaged)	Taking off
	21:54:00 ~ 22:04:00	1 (Disengaged)	Level flight
4	20:58:00 ~ 21:08:00	2 (Engaged)	Taking off
	23:25:00 ~ 23:35:00	1 (Disengaged)	Level flight

5.3 EEG Data Processing

The data processing procedure for the EEG sensors is shown in Figure 8. We start with removal of environmental and DC artifacts, then removal of EEG datasets with unreasonable measurements based on standard deviation (such as 0 indicating no signal collected), selection of EEG channels of interest, identification of spikes/excursions/amplifier saturation, removal of artifacts, calculation of Power Spectrum Density (PSD), and analysis with two different techniques, Fisher score and ANOVA, for engagement assessment.

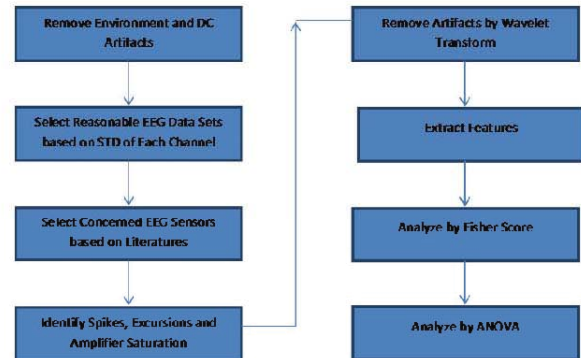


Figure 8. EEG data processing

ABM [20] suggested bi-polar sites, Fz-POz, Cz-POz for engagement assessment, and C3-C4, Cz-POz, F3-Cz, F3-C4, Fz-C3, Fz-POz for workload assessment. Leonard J. Trejo [21] emphasized that mental fatigue was associated with Fz, P7 and P8. To analyze pilots' mental states during the simulated flight, we started with all channels mentioned in these studies. The EEG sensors being used (actiCAP) did not

provide the channel POz and we selected Oz as a substitute, which is the nearest sensor to POz. To make it comparable, the sensor P7 and P8 were paired with Oz respectively. Finally, the selected EEG sensors were Fz-Oz, Cz-Oz, C3-C4, F3-Cz, F3-C4, Fz-C3, P7-Oz and P8-Oz.

In this study, EEG absolute PSD variables for each 1-s epoch were computed. For each bipolar pair, the power spectrum within each band was summed up as a feature. All the features were analyzed based on a Fisher score, which is the normalized distance between data points belonging to different states. The larger the Fisher score is, the farther the distance between different states is, indicating a better feature. By sorting the aforementioned features, we can find the most valuable features.

Furthermore, we analyzed the features using ANOVA. For example, P8-Oz in Gamma band was examined and its PSD in an engaged state (X label: 2) is significantly higher than that in a disengaged state (X label: 1), as shown in Figure 9. The analysis results confirm the usefulness of the features being extracted.

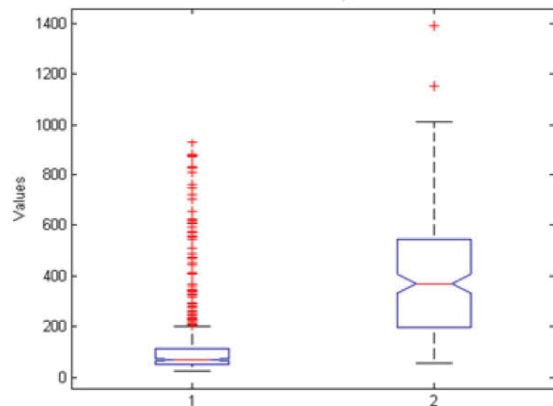


Figure 9. ANOVA for subject 2 (P8-Oz, 25-40Hz)

5.4 Eye Tracking Data Processing

Two types of eye movements have been studied: fixations and saccades with respect to attention allocation. Fixation is defined as a single point of gaze vector within a

threshold of two degrees for a minimum duration of 200ms [13].

Saccadic movement is derived by counting a saccade as the movement from one fixation to the next. Saccadic movements are measured by saccadic distance (deg). Their Euclidian distance can be derived by determining the plane on which the fixation is occurring and identifying the distance between that specified location and the eye gaze origin.

For an engaged pilot, the fixation duration is usually smaller than that of a disengaged pilot, who may be in a state of day dreaming or high fatigue. A disengaged pilot during the enroute phase may have longer fixation durations and/or increased saccade length due to decreased workload [13].

Figure 10 shows the fixation during the flight for subject 3. When failures happened around 20:00, we can see that failures are followed by a valley of fixation value, which implies that the engagement level increases when the pilot is faced with an emergent event (as shown in Figure 10). This verifies the feasibility of using fixation as an indicator for engagement evaluation. Similar observations can be made for saccades in the flight simulation.

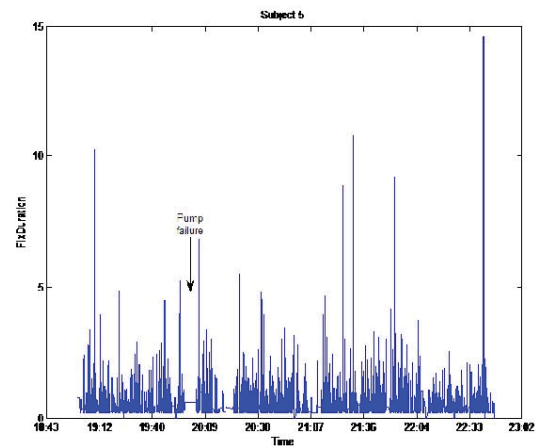


Figure 10. Examples of eye fixation duration

5.5 OFS Assessment Modeling Performance

We first evaluated the performance of the enhanced committee machine-based OFS assessment model using a 5-fold cross validation technique. More specifically, for each individual, we divided the whole dataset into five folds (equally) and trained an individual's model using four of them. The performance was evaluated by testing the model with the remaining fold. The performance results range from 97.2% to 99.8% for all four subjects.

We also built a model for each individual based on very limited training data, which contains the first X% of data samples for a subject in each engagement state (the value of X can be 5, 10, 15 and 20). The dataset was normalized by the mean and standard deviation of the extracted training data. The trained model was then tested with the remaining data from that subject. The evaluation results for the four individual models are shown in Figure 11. The detection accuracy for the OFS assessment of the four subjects with only 10% of data can reach 84.2%, 95.3%, 89.7%, and 99.4%, respectively.

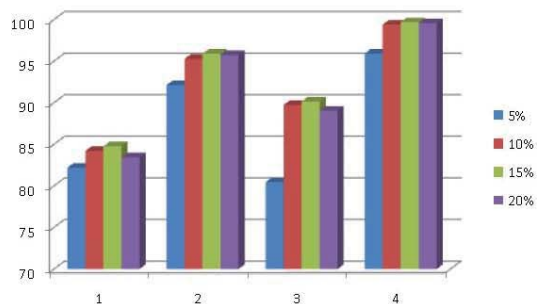


Figure 11. Performance evaluation results for four subjects.

6.0 DISCUSSIONS

In this research, we have successfully developed a systematic approach for engagement assessment. The approach is based on an understanding of the relationship between performance, workload, and engagement. Future tasks include enhancing the model with additional

sensory information (flight technical data and ECG, for example) and continuing to further verify and validate the real-time assessment technique with additional participants' data.

7.0 ACKNOWLEDGMENTS

This project was funded by the NASA (Contract No: NNX10CB27C). We thank Dr. Alan T. Pope, Mr. Chad L. Stephens, and Dr. Kara Latorella for their comments and suggestions as we performed this research.

8.0 REFERENCES

- [1]. G. Robert J. Hockey, Robert Hockey, "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," NATO ASI SERIES Vol 355
- [2]. W. Boucsein, R. W. Backs, "Engineering psychophysiology as a discipline: historical and theoretical aspects," in Boucsein, W., Backs, R.W. (Eds.), Engineering Psychophysiology: Issues and Applications. Erlbaum, London.
- [3]. Paul C. Schutte, et al., "The Naturalistic Flight Deck System: An Integrated System Concept for Improved Single-Pilot Operations," NASA report, December 2007
- [4]. D. Schmorow & A. Kruse, "Improving human performance through advanced cognitive system technology. Interservice/Industry Training," Simulation and Education Conference, Orlando, FL; 2002.
- [5]. D. Schmorow, K.M. Stanney, G. Wilson, and P. Young, "Augmented cognition in human-system interaction," In: Salvendy G, ed. Handbook of human factors and ergonomics, 3rd ed. New York: John Wiley; 2005.
- [6]. F. Karl, Van Orden, Erik Viirre, and David A. Kobus, "Augmenting Task-Centered Design with Operator State Assessment Technologies," Foundations of Augmented Cognition: Lecture Notes in Computer Science, 2007, Volume 4565/2007, 212-219.
- [7]. G. Zhang et al., "Individualized Cognitive Modeling for Close-Loop Task Mitigation," MODSIM World Conference, 2010, Virginia Beach, VA
- [8]. G. Zhang et al., "Individualized Cognitive Modeling for Closed-Loop Task Mitigation,

- ModSim World Conference, 2009, Virginia Beach, VA
- [9]. M.E. Smith, L. McEvoy, and A. Gevins, "Neurophysiological indices of strategy development and skill acquisition," *Cognitive Brain Research*, Vol. 7, issue 3, Jan. 1999, pp. 389– 404
 - [10]. Leonard J. Trejo et al., "EEG-based Estimation of Cognitive Fatigue," *Proceedings of SPIE: Bio-monitoring for physiological and cognitive performance during military operations*. v5797. 105-115.
 - [11]. R. H. Stevens , T. Galloway, and C. Berka, "EEG-Related Changes in Cognitive Workload, Engagement and Distraction as Students Acquire Problem Solving Skills," *Proceedings of the 11th international conference on User Modeling*
 - [12]. A. Pope, E. H. Bogart, D. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological Psychology*; 40(1): 187-195. (1995)
 - [13]. K. Ellis, "Eye Tracking Metrics for Workload Estimation in Flight Deck Operations, Masters Thesis, University of Iowa, 2009.
 - [14]. D. G. Jones, "Subjective Measures of Situation Awareness," In Endsley, M. R., & Garland, D. J., *In Situation Awareness Analysis and Measurement* (pp. 113-128). Mahwah: Lawrence Erlbaum Associates, Publishers.
 - [15]. V. J. Gawron, "Human Performance, Workload, and Situational Awareness Measures Handbook," Boca Raton: CRC Press Taylor & Francis Group.
 - [16]. S. W. Samn & L. P. Perelli, "Estimating aircrew fatigue: a technique with implications to airlift operations," Brooks AFB, TX: USAF School of Aerospace Medicine; 1982. Technical Report No 1982; SAM-TR-82–21
 - [17]. R. Farmer & N. D. Sundberg, "Boredom proneness: The development and correlates of a new scale," *Journal of Personality Assessment* 50 (1): 4–17.
 - [18]. J. Li, J. Yao, N. Petrick, R. M. Summers, and A. K. Hara, "Hybrid Committee Classifiers for a Computerized Colonic Polyp Detection System," *Medical Imaging 2006: Image Processing*, vol. 6144, pp. 1701-1709, 2006.
 - [19]. F. Li et al., "Engagement Assessment Using EEG signals," Accepted, MODSIM World Conference, 2011, Virginia Beach, VA
 - [20]. C. Berka, et al., "EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks," *Aviation, Space and Environmental Medicine*, Vol. 78, No. 5, Section II, May 2007
 - [21]. L. J. Trejo et al., "EEG-Based Estimation of Mental Fatigue: Convergent Evidence for a Three-State Model," *HCI International 2007 and Augmented Cognition International Conference*, Beijing, China, July 22-27. In D.D. Schmorow, L.M. Reeves(Eds.): *Augmented Cognition*, HCII 2007, LNAI 4565, pp. 201-211, New York: Springer LNCS.

3.6 Engagement Assessment Using EEG signals

Engagement Assessment Using EEG signals

Feng Li, Jiang Li, and Frederic McKenzie

Flix003@odu.edu; ili@odu.edu; rdmckenz@odu.edu

Guangfan Zhang, Wei Wang, Aaron Pepe, and Roger Xu
Intelligent Automation, Inc.

gzhang@i-a-i.com; jleddo@i-a-i.com; hqxu@i-a-i.com

Tom Schnell, Nick Anderson, Dean Heitkamp
University of Iowa

richeyc@ccad.uiowa.edu; thomas-schnell@uiowa.edu

Abstract. In this paper, we present methods to analyze and improve an EEG-based engagement assessment approach, consisting of data preprocessing, feature extraction and engagement state classification. During data preprocessing, spikes, baseline drift and saturation caused by recording devices in EEG signals are identified and eliminated, and a wavelet based method is utilized to remove ocular and muscular artifacts in the EEG recordings. In feature extraction, power spectrum densities with 1 Hz bin are calculated as features, and these features are analyzed using the Fisher score and the one way ANOVA method. In the classification step, a committee classifier is trained based on the extracted features to assess engagement status. Finally, experiment results showed that there exist significant differences in the extracted features among different subjects, and we have implemented a feature normalization procedure to mitigate the differences and significantly improved the engagement assessment performance.

1.0 INTRODUCTION

Operator Functional State (OFS) assessment is important in aviation to ensure mission success and improve mission performance. According to North Atlantic Treaty Organization (NATO) [1], OFS is defined as the multidimensional pattern of human psychophysiological condition that mediates performance in relation to physiological and psychological costs. The application of EEG signals for human mental states assessment, such as mental fatigue, operator engagement, and workload, has been widely studied.

In this paper, we present methods to analyze and improve an EEG-based engagement assessment approach. The developed methods were evaluated with EEG signals collected from four pilots. More specifically, the raw EEG recordings were first cleaned up by removing spikes, baseline drift and other artifacts in a data preprocessing step. We then computed 1-Hz bin PSDs from 1 Hz to 40 Hz as features. Those extracted features were subsequently fed into a committee machine classifier to identify if the pilots were engaged. To train the committee classifier,

engagement ground truth for each subject was identified using the methods described in the accompanying paper [11].

We also presented feature analysis results using the Fisher score measure and the one-way ANOVA method. Fisher score is a measure describing how good a feature is for differentiating one category from another, i.e., engaged vs. disengaged, while the ANOVA method performs hypothesis tests to verify if features coming from different categories are statistically different. We found that features identified for each subject by Fisher score can effectively discriminate different engagement states for the subject. However, significant differences in EEG signal features among different subjects were found by the ANOVA method.

To mitigate the feature differences among subjects, we performed a normalization procedure on the extracted features for each subject to be tested. We assume that there was a small set of EEG recordings available for the subject, i.e., from a baseline experiment [2]. The means and standard deviations were then computed for features from the small available dataset

and were subsequently used to normalize all remaining data for that subject. Finally, we evaluated the proposed methods for engagement assessment on the data of four subjects using a 5-fold cross validation method.

The remainder of the paper is organized as follows: section 2 presents our methods to assess the pilot's engagement state. Section 3 illustrates the achieved results and section 4 concludes this paper.

2.0 METHODS

2.1 Data collection

In order to study engagement, experiments in a fully equipped Boeing 737 simulator were conducted [11]. Participants involved pilots with commercial/private/ATP (Airline Transport Pilot) licenses. Each experiment simulated a flight from Seattle Tacoma International Airport to Chicago O'Hare International Airport. Video, audio and physiological information, including EEG, ECG and eye tracking were recorded during the whole simulated flight.

2.2 Engagement ground truth finding

Although there was no sensor to provide online engagement ground truth, the ground truth can be assessed by incorporating pilots' self-evaluation and behavioral measurements etc. as we described in the accompanying paper [11]. For example, in the phase of taking off, landing or a simulated failure, a pilot was more likely to be engaged, while in the phase of flight or the pilot was napping, it is likely that he was disengaged. By using the method described in [11], we identified several data segments for the "engaged" and "disengaged" states as listed in Table 1 in this study to validate the proposed methods.

2.3 EEG artifact removal

Although it has been found that EEG signals from certain sensor locations and EEG features in different frequency bands are highly correlated to OFS [1]-[4], EEG

recordings are known to be often

Table 1: Pilots' engagement states (1: disengaged; 2: engaged)

Subject	Time duration	State	comments
1	19:08 ~ 19:18	2	taking off
	21:08 ~ 21:17	1	flat flight
2	19:52 ~ 20:03	2	taking off
	21:19 ~ 21:29	1	flat flight
3	19:13 ~ 19:23	2	taking off
	21:54 ~ 22:04	1	flat flight
4	20:58 ~ 21:08	2	taking off
	23:25 ~ 23:35	1	flat flight

contaminated by physiological artifacts from various sources, such as eye blinking/movement, heart beating and movement of other muscle groups [5]. Artifacts are often mixed together with brain signals, making interpretation of EEG signals difficult [6]. To perform OFS assessment using EEG recordings, it is critical to exclude EEG artifacts contained in the signals. Many methods, such as Principle Component Analysis (PCA) [7] and Independent Component Analysis (ICA) [8]-[9] have been developed to remove the EEG artifacts. Wavelet-based methods have also received significant attentions [10] for EEG artifact removal.

In this paper, the proposed procedure for artifact removal is shown in Fig. 1. We started with the removal of environmental artifacts by applying a 60-Hz notch filter followed by the removal of the baseline drift by utilizing a high-pass filter with a cutoff frequency of 0.5-Hz. Due to subject movements in the experiments, some EEG sensors failed because of loose connections between the sensors and scalp for a specific time duration. To remove signals from those sensors, we checked the standard deviations of each EEG channels, and those with zero or values much larger than the average standard deviation, computed from other valid channels, were

discarded. According to the criteria used in the ABM's model [2], EEG segments were identified as spikes or baseline drifts if the EEG amplitude changed abruptly (100uv over a short duration of 30 ms in this paper).

Once the environmental and baseline drifts were removed, the remaining EEG signals were segmented to epochs using a 3-second window. Our EEG signals were sampled at the frequency of 200 Hz and each epoch contains 600 data-points. In our study, two consecutive epochs had 2 seconds or 400 data points of overlapping. To remove ocular and muscular artifacts, each epoch was decomposed using a six level's stationary wavelet transformation, yielding a set of wavelet bands: 0-1.56, 1.56-3.13, 3.13-6.25, 6.25-12.5, 12.5-25, 25-50, and 50-100 Hz. For each wavelet band, the mean and standard deviation of the coefficients were calculated. Coefficients in the band were set to its mean if the absolute difference between the coefficient and the mean was larger than 1.5 times of the standard deviation in that band. Finally, the EEG signals were reconstructed from the modified coefficients.

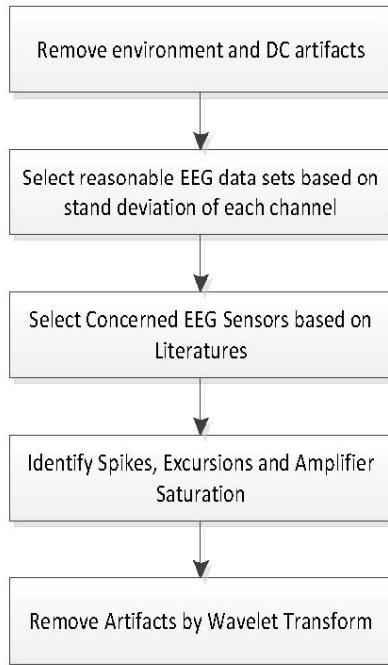


Figure 1: EEG signal preprocessing

2.4 Feature extraction

For each decontaminated epoch, 1-Hz bins of Power Spectral Densities (PSD) from 1 Hz to 40 Hz were calculated as features. In our study, we utilized eight channels of EEG recordings, which yielded 312 (8*39) features for each epoch.

2.5 Feature analysis

We employed two methods, Fisher score and one-way ANOVA, to analyze the extracted features for the subjects. Fisher score was calculated based on the following formula: $\frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}$ where $m_{1,2}$ and $\sigma_{1,2}$ are the means and standard deviations of the data points belonging to state 1 or 2 (engagement or disengagement), respectively. The Fisher score is the normalized distance between data points belonging to different states. The larger the fisher score, the more powerful the feature is. By sorting features based on the Fisher score, we can identify the most effective ones that can differentiate the two states. On the other hand, the one-way ANOVA analysis provides an intuitive way to compare data points belonging to difference groups/states. This method can help us verify the effectiveness of those features ranked by the Fisher score for each individual or across subjects.

2.6 Committee machine classifier

We utilized a committee machine as the classification model, which was developed previously [12]. A committee machine is an ensemble classifier consisting of multiple classifiers whose responses are combined as a single response. Fig. 2 shows the main procedures:

- Use the bootstrapping technique to 'disturb' the training data, resulting different sets of training data,
- Train a Multi-layer perceptron (MLP) on each set of training data. The trained model is regarded as a base classification model/committee member.

To , to make each of the committee members diversified, we apply an advanced feature selection algorithm, Piecewise Linear Orthogonal Floating Search (PLOFS) [13],

- Delete the committee members having high biases (accuracy < 50%),
- Utilize the majority vote scheme to fuse decisions from committee members.

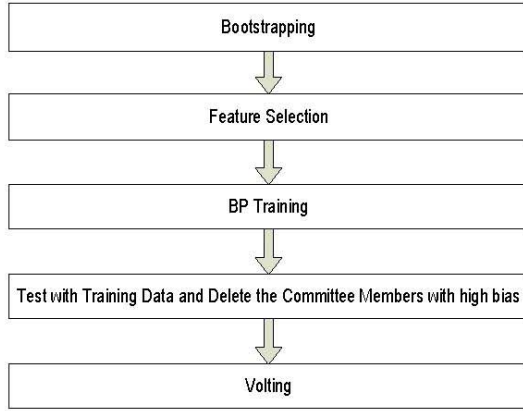


Figure 2: Diagram of the committee machine

2.7 Baseline normalization

It has been observed that there were significant differences in the PSD features among different subjects. To address the individual variation, we assume that a small set of data samples from the test subject are available (i.e., from baseline experiments), and normalize the features based on the small dataset from the test subject. Mean and standard deviation of the subject were computed from his/her available data samples and the remaining data from the subject was normalized by the computed mean and standard deviation.

2.8 Diagram of the proposed approach

Combining the components described above, we present an integrated approach as shown in Fig. 3, where we assume that our subjects are available for the assessment. The purpose of the system is to adapt the model trained on subjects 1, 2 and 3 to subject 4.

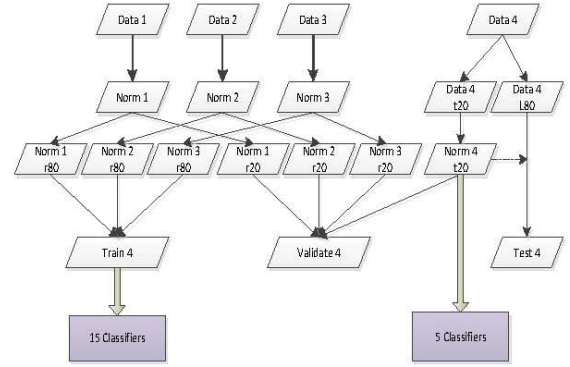


Figure 3: System diagram for the committee machine based engagement assessment

1. Data from subjects 1, 2 and 3 are normalized by their own mean and standard deviations,
2. For the data set from subject 4, top 20% data samples are extracted and normalized by their own mean and standard deviation, forming dataset 'Norm4_t20'. The mean and standard deviation are then utilized to normalize the remaining 80% of the data points,
3. From the three normalized data sets from subjects 1, 2 and 3, 80% of them are extracted and combined as the training dataset (train4); the other 20% are combined with the top 20% of data samples from subject 4 as the validation dataset (validate 4),
4. 15 classifiers are trained based on 'Train4'; 5 other classifiers are trained based on 'Norm4_t20',
5. 20 classifiers are applied to 'Validate 4' and only those classifiers with accuracy over 50% will be kept for testing,
6. The kept classifiers are then utilized to form the final classification results by majority voting.

3.0 RESULTS

3.1 Data preprocessing

Fig. 4 (a) shows a raw segment recording in which baseline drift and other artifacts are presented. Fig. 4 (b) shows the signal after the drift being removed and Fig. 4 (c) shows the "clean" data after the ocular artifact

being eliminated. It is observed that the proposed methods can effectively remove

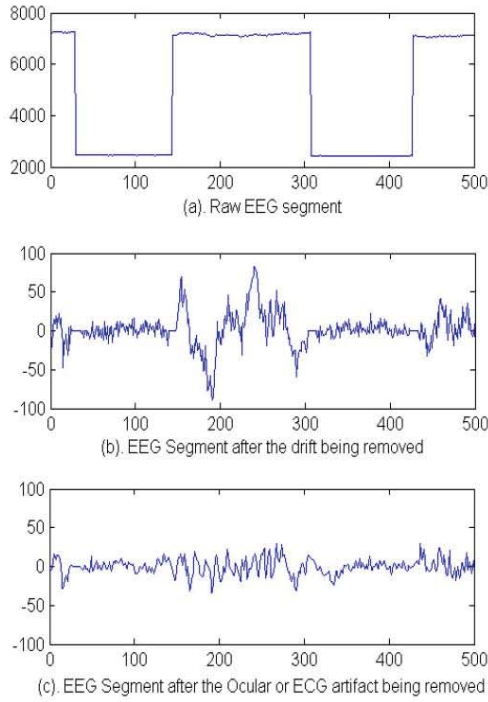


Figure 4: Artifacts removal

common artifacts contained in raw EEG recordings.

3.2 Feature analysis by Fisher score

To find the most valuable features, we firstly calculated the Fisher score for each feature of each subject. Then features were sorted and finally we selected the common features from the top 150 features for the four subjects. In our study, we found 42 common features as shown in Table 2. Four bi-polar sites, Cz-Oz, P7-Oz, P8-Oz, and Pz-Oz were involved and most of the features were in the frequency range of 25~40 Hz.

Table 3 shows top 10 common features and it is clear that the 39 Hz PSD bin from Cz-Oz is the most valuable feature. Figures 5-8 show one common feature among the four subjects, where the feature is large in magnitude for the 'engaged' state and small for the 'disengaged' state. It is also

observed that features from different subjects are not in the same scale, implying that the data from different subjects cannot be combined directly and necessary adaptation techniques are mandatory. In our study, we normalized each feature from a subject by its own mean and standard deviation computed using a small dataset from the subject.

Table 2. Distribution of high rank features

	1~4 Hz	5~7 Hz	8~13 Hz	14~24 Hz	25~40 Hz
Cz-Oz	0	0	0	0	8
P7-Oz	4	1	0	9	11
P8-Oz	0	0	0	1	7
Pz-Oz	0	0	0	0	1

Table 3: Fisher scores for the top 10 common features

Subject 1	Subject 2	Subject 3	Subject 4	Sensor	band
1.54	1.61	0.39	4.32	Cz-Oz	39
0.31	2.27	0.57	1.41	P7-Oz	35
0.22	1.78	0.69	1.60	P7-Oz	32
0.29	1.67	0.60	1.59	P7-Oz	29
1.19	1.50	0.35	0.94	Cz-Oz	38
0.30	1.81	0.56	1.29	P7-Oz	34
0.25	1.59	0.65	1.45	P7-Oz	33
0.77	1.90	0.26	0.99	Pz-Oz	33
0.32	1.61	0.52	1.45	P7-Oz	31
0.29	1.60	0.68	1.34	P7-Oz	28

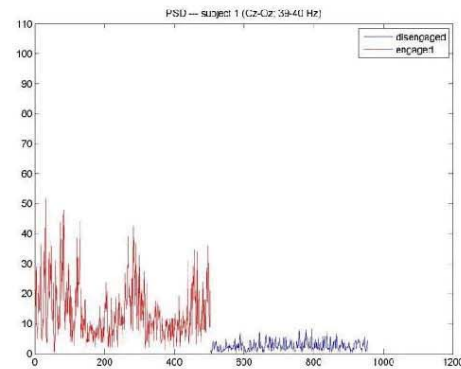


Figure 5: Feature Cz-Oz from 39-40 Hz for subject 1

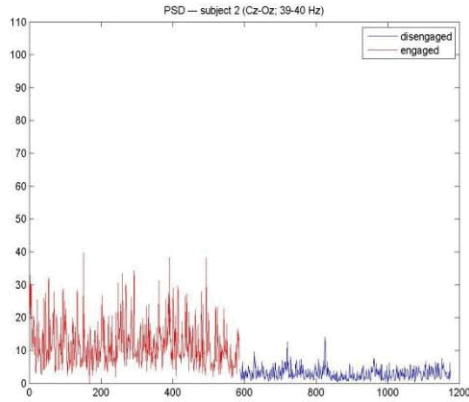


Figure 6: Feature Cz-Oz from 39-40 Hz for subject 2

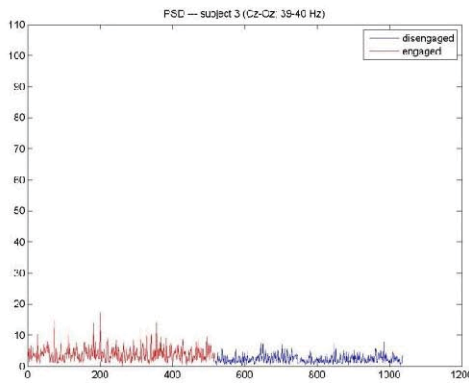


Figure 7: Feature Cz-Oz from 39-40 Hz for subject 3

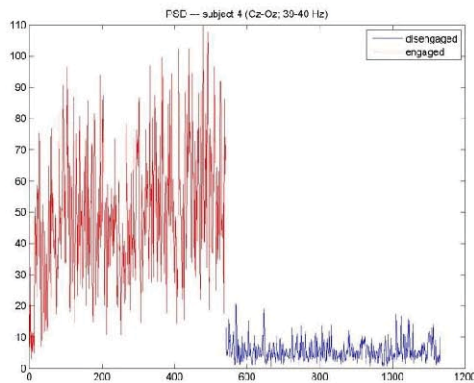


Figure 8: Feature Cz-Oz from 39-40 Hz for subject 4

3.3 Feature analysis by ANOVA

We also analyzed the features using the one-way ANOVA tool and results confirmed that for the feature ranked first by the Fisher score, its PSD for the engaged state is significantly larger than the PSD for the disengaged state. Fig. 9 shows the result of ANOVA analysis for subject 1 and results for other subjects are similar. Fig. 10 shows the ANOVA analysis for the four subjects all together, where the features for each subject were paired. It is clear that features from different subjects are significantly different even if they were all engaged.

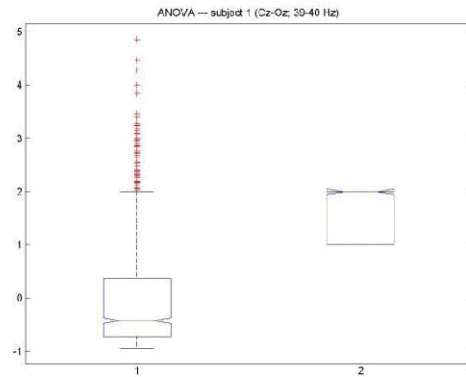


Figure 9: ANOVA analysis for subject 1

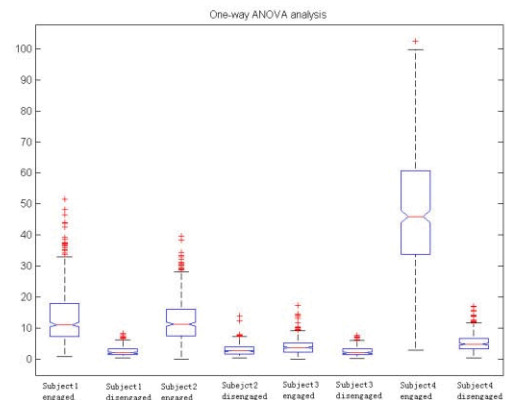


Figure 10: ANOVA analysis for four subjects

3.4 Classification results

We studied two scenarios of the individual assessment model, in which training and testing of the committee machine was performed for each subject independently and results are shown in Table 4. In scenario 1, we used a 5-fold cross-validation scheme to evaluate the engagement assessment model and the average accuracy is 98.55%. In scenario 2, we trained the model by the top 20% data samples from each subject and tested the model on the remaining data. The average accuracy is 94.01%. Both of them prove the effectiveness of the extracted features.

In addition, two scenarios had been studied for evaluating the proposed system if data samples from a subject are not available or very limited. For each of the four subjects, we first trained an average model by using the data from the other three subjects. In scenario 1, we assumed that data samples from the subject to be tested were not available and we normalized the testing data by the mean and standard deviation computed from other subjects, producing poor accuracies as shown in Table 5. In scenario 2, we assumed that a small set of data samples from the subject to be tested were available (the top 20% of the testing data) and used their mean and standard deviation to normalize the remaining testing data. It can be observed in Table 5 that the results are significantly improved.

Table 4: Classification accuracy (%) of the individual models

Scenario	subject 1	subject 2	subject 3	subject 4	Average
1	98.76	97.18	98.43	99.82	98.55
2	90.35	95.32	90.49	99.89	94.01

Table 5: Classification accuracy (%) of the average models

Scenario	Subject 1	Subject 2	Subject 3	Subject 4	Average
1	52.66	78.55	53.95	97.26	70.60
2	93.21	91.66	89.11	69.88	85.96

4.0 CONCLUSIONS

In this research, we proposed methods for engagement assessment based on EEG signals. The methods include EEG artifacts removal, feature analysis and ranking, and a feature normalization procedure. Experimental results illustrated that the artifact removal methods eliminated most of the artifacts in the EEG recordings. Feature analysis showed that engagement can be effectively assessed using the identified features but there existed large differences in features from different subjects. Finally, we demonstrated that the feature normalization procedure significantly mitigated feature variations across subjects.

5.0 ACKNOWLEDGMENTS

The OFS assessment model development is being funded by the NASA (Contract No: NNX10CB27C). We thank Dr. Alan T. Pope, Mr. Chad L. Stephens, and Dr. Kara Latorella, for their comments and suggestions as we performed this research.

6.0 REFERENCES

- [1]. G. Robert J. Hockey, Robert Hockey, "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," *NATO ASI SERIES* vol. 355.
- [2]. Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, Patrick L. Craven, "EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks," *Aviation, Space and Environmental Medicine*, vol. 78, No. 5, Section II, May 2007.
- [3]. Leonard J. Trejo, Kevin Knuth, Raquel Prado, Roman Rosipal, Karla Kubitz, Rebekah Kochavi, Bryan Matthews, Yuzhen Zhang, "EEG-Based Estimation of Mental Fatigue: Convergent Evidence for a Three-State Model," *HCI International 2007 and Augmented Cognition International Conference*, Beijing, China, July 22-27. In D.D. Schmorow, L.M. Reeves(Eds.):

- Augmented Cognition, HCII 2007, LNAI 4565, pp. 201-211, New York: Springer LNCS, 2007.
- [4]. Alan T. Pope, Edward H. Bogart, Debbie S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological Psychology*, vol. 40, pp. 187 – 195, 1995.
 - [5]. T. Jung, S. Makeig, C. Humphries, et. al., "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163-178, 2000.
 - [6]. E. Urretarazu, J. Iriarte, M. Alegre, M. Valencia, C. Vireri, and J. Artieda, "Independent component analysis removing artifacts in ictal recordings," *Epilepsia*, vol. 45, pp. 1071-1078, 2004.
 - [7]. P. Berg and M. Scherg, "A multiple source approach to the correction of eye artifacts," *Electroencephalography and clinical Neurophysiology*, vol. 90, no. 3, pp. 229-241, 1994.
 - [8]. P. Common, "Independent Component Analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994.
 - [9]. A. Delorme, S. Makeig and T. Sejnowski, "Automatic artifact rejection for EEG data using high-order statistics and independent component analysis," *Proceedings of the Third International ICA Conference*, Dec 9-12, San Diego, 2001.
 - [10]. V. Krishnaveni, S. Jayaraman, L. Anitha and K. Ramadoss, "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients," *Journal of Neural Engineering*, vol. 3, pp. 338-346, 2006.
 - [11]. Guangfan Zhang, Wei Wang, Aaron Pepe, Roger Xu, Tom Schnell, Nick Anderson, Dean Heitkamp, Jiang Li, Feng Li and Frederic McKenzie, "A Systematic Approach for Real-time Operator Functional State Assessment," *ModSim 2011*.
 - [12]. Feng Li, Frederick McKenzie, Jiang Li, Guangfan Zhang, Roger Xu, Carl Richey, Tom Schnell, "Imbalanced Learning for Functional State Assessment," *MODSIM 2010*.
 - [13]. Jiang Li et al., "Feature Selection Using a Piecewise Linear Network," *IEEE Trans. Neural Network*, vol. 17, no. 5, pp. 1101-1115, 2006.

3.7 Modern Warfare: An M&S Examination of the Dynamic Impact of Warlords and Insurgents on State Stability

Modern Warfare: An M&S Examination of the Dynamic Impact of Warlords and Insurgents on State Stability

Phillip N. Jones
MYMIC, LLC.

phillip.jones@mymic.net

Nick Drucker
MYMIC, LLC.

nick.drucker@mymic.net

Noah Schwartz, Ph.D.
Christopher Newport University
noah.schwartz@cnu.edu

Abstract: 9/11 changed the world as we knew it. Part of this change was to redirect the military of the United States away from focusing primarily on conventional conflict to a primary focus on unconventional or irregular conflict. This change required a tremendous learning effort by the military and their supporting research and development community. This learning effort included relearning of old but largely forgotten lessons as well as acquiring newly discovered knowledge. During the process of our immediate 9/11 response, we identified that we were engaged in Iraq and Afghanistan in an insurgency. Subsequently, our focus converged upon the description of insurgencies and the requirements for counterinsurgency. This paper argues that emerging conditions now allow the re-evaluation of the type of conflict occurring today and into the foreseeable future: that we, including the modeling and simulation world, emerge from a singular focus on orthodox insurgencies and start to consider the consequences and opportunities of the complexity of current conflicts. As an example of complexity, this paper will use the relatively common phenomenon of the Warlord or Warlordism. The paper will provide a definition of this phenomenon and then describe the implications for modelers. The paper will conclude by demonstrating the impact of incorporating this one rather prosaic complexity into an insurgency model, using agent based modeling (ABM).

1. INTRODUCTION

9/11 changed the World as we knew it. Part of this change was to redirect the military of the United States away from an almost exclusive focus on conventional conflict to a primary focus on unconventional or irregular conflict. This change required a tremendous learning effort by the military. The supporting research and development community was part of this learning process. This effort included relearning old but largely forgotten lessons as well as acquiring newly discovered knowledge.

The exigencies of our response to 9/11, largely wrapped up in Operations Enduring and Iraqi Freedom, and our general unpreparedness for the unconventional

conflicts emerging from these operations, required a strategically rushed response. Clausewitz observed that the first requirement of war is to identify the form of war you are fighting. During the process of our immediate 9/11 response, we identified that, in Afghanistan and eventually in Iraq, we were fighting an insurgency. Subsequently, our focus converged upon the description of insurgencies and the requirements for counterinsurgency, or COIN. Even though security intellectuals are debating the pertinence of terms such as guerilla war, hybrid war, fourth generation war, unrestricted war, new war, etc to describe our current conflicts, operators have moved forward in labeling the current conflicts as insurgencies and our response, necessarily, as counterinsurgency, or COIN. This

deduction was solidified by the publication of the US Army and Marine Corps doctrine manual on Counterinsurgency: FM 3-24/MCWP 3-33.5 [1].

FM 3-24 borrows from Joint Publication 1-02 in defining insurgency as: "...an organized, protracted politico-military struggle designed to weaken the control and legitimacy of an established government, occupying power, or other political authority while increasing insurgent control" [2]. The same manuals provide a highly relative definition of COIN: "Those military, paramilitary, political, economic, psychological, and civic actions taken by a government to defeat insurgency" [2]. There is a level of monolithism in these definitions centered on the strategic goals of an insurgency: the transfer of political power from one group to another. This monolithism is also found in the common works on insurgency, including older works such as Galula's *Counterinsurgency Warfare* and Trinquier's *Modern Warfare*, and newer works such as Smith's *The Utility of Force* and Kilcullen's *Accidental Guerrilla*.

This paper argues that emerging conditions now allow the re-evaluation of the type of conflict occurring today and into the foreseeable future: that we, including the modeling and simulation world, emerge from a singular focus on orthodox insurgencies and start to consider the consequences and opportunities of the complexity of current conflicts. This first requires an appreciation of the complexity of the current conflict environment and its inherent and potential complexity and an appreciation of the implication of that complexity.

As an example of this complexity, this paper will use the relatively common phenomenon of the Warlord or Warlordism. Warlords arise and thrive in the power vacuum of weak and failed states—just the type of conditions into which the future will take the US military [see 6]. The paper will provide a definition of this phenomenon and then describe the implications for modelers. The paper will conclude by demonstrating the impact of incorporating this one rather

routine complexity into an insurgency model, using agent based modeling (ABM).

2. Describing the Warlord

Warlords are a common historical phenomenon. The specific term was coined to label local, militaristic leaders that dominated China between the collapse of the Ming Empire and the rise of the KMG [3]. However, all continents have seen Warlords. They existed in Europe through the Dark and later Ages following the collapse of the Roman Empire—Warlords and their retainers built many of the castles seen today in Europe. In North America, Comanche war party leaders represented Warlord traits [4]. Warlords existed in historical Japan. Today, Warlords are found in Africa and Southwest Asia, specifically in Afghanistan.

The word "Warlord" is a term applied by humans to label a particular social phenomenon. As with much from the social sciences, this is not a singularly discreet phenomenon with easily identifiable boundaries. Rather, it is location upon the vast spectrum of how humans organize themselves and, in the process, deal with other humans [3][4]. Thus, the definition and description of a Warlord and of Warlordism will always contain flexibility relative to the experiences and attitudes of the word's user. In the case of the term Warlord, an example of this flexibility may be found in the aspect as to whether the Warlord is or is not financially motivated [4].

This paper will use a broad definition of Warlord. A Warlord is an individual beholden to an external physical, intellectual, or emotional authority such as a state or cause; successful in leadership through charisma and other motivational qualities; possessing military organizational traits; and himself internally motivated by personal gain, be that gain physical (i.e. financial) or emotional (i.e. glory, reputation, etc).

This definition includes those elements that make the Warlord both similar

and dissimilar from the insurgent leader. The Warlord's personality and military organization allow him to collect and organize an effective force of retainers, personnel loyal first and, perhaps, only to him. This could be said of many leaders, including insurgent leaders. However, the insurgent leader—Mao, Begin, Castro, Pol Pot, Noriega, Zargawi—are in word if not in action, subservient to some higher calling, such as communism, nationalism, or fundamentalism. Through their insurgency, they seek initial instability in order to weaken the state so they can subsequently replace it, whereupon the insurgent imposes its own, personally motivated stability. The Warlord is interested in only his own gain. He seeks instability to create the conditions in which he thrives. He has no interest in transitioning to some new, stable situation. Thus, in the initial phases of an insurgency, the Warlord will work with the insurgent if not disguise himself as one—the Warlord's interest parallel the insurgent's. Later in the insurgency, when the winning insurgent begins to impose his own stability upon the state, the Warlord will turn on the insurgent, continuing to destabilize the situation. The example of this is Charles Taylor who entered Sierra Leone allied with the anti-government revolutionaries of the Revolutionary United Front (RUF) to start an insurgency, and then killed off those revolutionaries to seize control of RUF for his personal profit [3].

3. Methods and Model Construction

In order to experiment with the concepts presented above we constructed an agent based model (ABM) of insurgent and warlord instability activity. We constructed the model for this experiment using NetLogo [7], a software which is ideally suited for this type of modeling and simulation work. Within NetLogo it is possible to create large groups of agents, and by assigning them a set of rules to follow observe the aggregate results of each one's individual actions.

The model consisted of a population of 1,000 agents with varying levels of initial instability. The model also included a state security force trying to maintain stability and support for the current ruling group, an insurgent influence attempting to create instability and subsequently win populace support for their views, and a Warlord influence seeking to create and maintain instability. In order to test our hypothesis that Warlord influence will result in greater instability by complicating the re-stabilization process, we included a switch to allow the model to create a baseline by running without the Warlord influence and then run with the Warlord influence to measure the significance.

The model consisted of five variables of interest: the insurgent influence level, the security force influence level, the warlord influence level, the initial instability within the model, and the threshold for agents to become unstable. The influence levels represented the amount an agent's instability variable can change based on interaction with the security, insurgent, or warlord factions. The greater the influence the more the faction could change the agent's stability level. The initial instability within the model allowed us to seed the simulation with some level of instability. This allowed us to experiment with relatively stable situations or those already in a state of relative instability. Finally, the instability threshold allowed us to experiment with how resilient an agent is to becoming unstable. The higher the threshold the more negative influence an agent needed to become unstable.

During each run of the model the following procedures occurred: populace members received influence from the security force, the insurgent force acted to influence agents, if warlords are present they acted to influence agents, agents calculated their instability, the model calculated the overall instability. During the security force interaction, populace agents had a 15% chance of receiving influence from the security force. Because the force's goal was to maintain stability, they operated

throughout the entire system. However, to ensure they were not omniscient and thus able to constantly positively influence all agents, the 15% level meant on each run roughly 15% of the entire populace, or 150 agents, would receive positive influence. The insurgent would always reach a group of agents. However, which group they reached was random. This allowed us to mimic the way insurgent groups operate, not always attacking the same target day after day but selecting and attacking vulnerable targets. In the model, Warlords operated in a similar manner, always interacting with a random group of populace members. The main difference between the groups was that Warlords would only instigate instability but the insurgent would instigate instability and try to win support for their cause. During the instability check procedure, all agents calculated whether they had crossed the threshold to become unstable. The number of agents who were unstable was then fed to a master variable that determined when 60% of the populace had reached instability. Once this occurred, the model considered the populace unstable and noted the time. This also triggered a change in behavior in the model, with initial instability resulting in an attempt by insurgents to return order and support for their goals, while warlords simply continued to seek instability. Once this process began the model would note when a new stable rate occurred, if it ever occurred. The second instance of stability represents a shift by the populace to support of the insurgent. If 60% of the populace never reach instability, the model would never become unstable, and likewise if the model became unstable but never reached a second instance of stability this was noted.

To support our contention that the community expand its perspective beyond the traditional insurgency conflict model, we wanted to illustrate the impact of adding complexity to that model, specifically adding the rather common phenomenon of Warlordism. We thus created an experiment. We intentionally kept the experiment simple. The design required

creating a simple insurgency model within an ABM. We based this model on a generic insurgency scenario. The scenario postulated a small group disaffected against the government, i.e. the insurgents. That group desired to replace the government. To do so, they first had to remove the link between the existing government and the populace, i.e. create instability. Once instability reached a critical mass, the group then attached the population to themselves, i.e. create insurgent sponsored stability.

The design included creating a baseline. This was a scenario without a Warlord presence. The next part of the design was to include a Warlord and then determine if there was an impact, measure the impact, determine what that impact was, and measure the significance of the impact. The Warlord sought instability. While the insurgent focused on causing instability, the Warlord served as his ally. However, once the model reached the tipping point and the insurgent started to focus on causing stability, the Warlord would work against the insurgent. The experiment anticipated that in the initial phase, the Warlord would serve as an accelerant. In the second phase, the Warlord would serve as a modulator. To keep the experiment simple, it did not include any outside COIN force, i.e. an intervening United States. It did, however, include a local security force. This force worked to maintain stability. The local security force acted upon the entire population but with reduced effectiveness due to being spread out. Additionally, the security force lost influence near the tipping point when the population became unstable.

3.1. Experimental Protocol

We conducted 17,280 runs of the model under varying conditions. The main purpose of the experiment was to measure the impact of adding a Warlord to a conflict by determining if the impact of the inclusion on the speed that a population would both become initially unstable and subsequently re-stabilizing under the insurgent influence. While this was the main thrust of the

experiment, we also examined the impact upon results of the variables governing insurgent influence, Warlord influence, security influence, instability thresholds, and initial instability within the model.

4. RESULTS

The results of the experiment demonstrate the significance of adding complexity into an insurgent model. The inclusion of a Warlord presence in the model supported our hypothesis that including the Warlord presence would result in faster times to initial instability but a much longer run to reach a new level of stability. Tables 1 and 2 below outline the number of steps it took the model to reach instability and then re-stabilize (ticks), the number of turns it took the model to reach initial instability (un-stable-turn), and the number of times out of 8,640 trials that the model reached a level of initial instability. As the tables show, when the Warlord was not present, it took the model almost 300 additional turns to complete. Completion meaning the model reached an initial unstable level and then returned to stability. In addition, it took over 40 more turns to reach the initial instability. These numbers represent a 36% increase in steps to complete and 26% increase in time taken to reach initial instability when there is no Warlord presence. The number of unstable trials also demonstrates that Warlord presence greatly increased the likelihood of instability as there was a 59% increase in trials where instability occurred when the warlord was present.

Table 1: No Warlord present

Warlord Off	
ticks	un-stable-turn
803.75	177.06
Number of trials unstable	
1928	

Table 2: Warlord present in the model

Warlord On	
ticks	un-stable-turn
512.89	130.49
Number of trials unstable	
4747	

In order to determine if the results we observed in these experiments were significant we conducted an Independent Samples T-Test on the data. The results of the test appear in Table 3 and demonstrate that there was a significant difference between the two sets of experiments. This confirms our observation that including a Warlord in the model will cause quicker time to initial instability (p-value < .01), but result in a longer amount of time until a new level of stability occurs (p-value < .01).

Table 3: Two Sample T-Test

Group Statistics					
	WarLord	N	Mean	Std. Deviation	Std. Error Mean
Ticks	.00	8640	803.7503	423.77172	4.55906
	1.00	8640	512.8897	421.80117	4.53786
UnstableTurn	.00	8640	39.5109	109.62365	1.17936
	1.00	8640	71.6947	131.61067	1.41591

Independent Samples Test					
		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Ticks	Equal variances assumed	11.919	.001	45.217	17278
	Equal variances not assumed			45.217	17277.625
UnstableTurn	Equal variances assumed	235.898	.000	-17.465	17278
	Equal variances not assumed			-17.465	16731.166

Independent Samples Test					
		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	
Ticks	Equal variances assumed	.000	290.86065	6.43251	
	Equal variances not assumed	.000	290.86065	6.43251	
UnstableTurn	Equal variances assumed	.000	-32.18380	1.84274	
	Equal variances not assumed	.000	-32.18380	1.84274	

Independent Samples Test					
		t-test for Equality of Means			
		95% Confidence Interval of the Difference			
		Lower		Upper	
Ticks	Equal variances assumed	278.25228		303.46902	
	Equal variances not assumed	278.25228		303.46902	
UnstableTurn	Equal variances assumed	-35.79575		-28.57184	
	Equal variances not assumed	-35.79576		-28.57183	

In addition to the findings regarding the presence of Warlord impact on instability, we examined how the other variables related to the time taken to reach instability. In order to determine what relationships existed and if they matched our hypothesis', we employed regression analysis. Our hypothesis was that insurgent influence, stabile influence, initial instability, and the warlord influence would all have negative relationships with the time taken to reach instability. By this we mean that where these variables increase, time to instability would likely decrease. In addition, we included the instability threshold in the regression, hypothesizing that the higher the threshold the longer it would take the

model to reach instability. Tables 3 and 4 below display the regression statistics.

Table 4: Regression without Warlord

<i>Regression</i>	
Multiple R	0.61
R Square	0.37
Adjusted R2	0.37
Stand Error	104.14
Trials	8640

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P</i>
Intercept	157.72	5.59	28.21	0.00
ins-inf	-43.48	0.67	-64.97	0.00
stab-inf	32.01	0.71	44.83	0.00
initial	-0.63	0.07	-9.01	0.00
war-inf	-7.03	0.53	-13.20	0.00
threshold	1.59	0.07	23.15	0.00

Table 5: Regression with Warlord

<i>Regression</i>	
Multiple R	0.61
R Square	0.37
Adjusted R2	0.37
Stand Error	104.14
Trials	8640

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P</i>
Intercept	157.72	5.59	28.21	0.00
ins-inf	-43.48	0.67	-64.97	0.00
stab-inf	32.01	0.71	44.83	0.00
initial	-0.63	0.07	-9.01	0.00
war-inf	-7.03	0.53	-13.20	0.00
threshold	1.59	0.07	23.15	0.00

For almost all of our hypotheses, we were able to confirm our initial projections. However, we did discover one interesting note. When the Warlord presence did not exist, the relationships between insurgent influence and stable influence were opposite of what we expected, i.e. in these runs, an increase in insurgent influence would result in a longer time to instability and increased stable influence would result in quicker time to instability. This may be a result of the way we ran our experiment. Due to computational limitations, we choose only to use a strong security force and a

weak security force. Because we did not sample across more variable values, this may be a demonstration of sensitivity within the model. In future runs, we would experiment across more variable values and examine whether the results remain the same. Despite this finding, the model appears to function properly and we were able to confirm our main hypothesis that the presence of a Warlord influence will result in quicker time to initial instability, but a prolonged time to re-stabilize.

5. CONCLUSION

The results of this experiment support proposal of the requirement to move beyond simple applications and incorporate complexity in our appreciation of contemporary conflict. Our experiment confirmed our initial hypothesis that the inclusion of a complicating factor, such as a Warlord, will significantly affect an insurgency model. This research does not approach addressing all of the variables that would inevitably be present in an actual insurgency. Thus, it is not a beginning to better understanding but, hopefully, a beginning to a beginning.

6. REFERENCES

- [1]. FM 3-24, (2009), *Field Manual 3-24, Tactics in Counterinsurgency*, Headquarters, United States Army, Washington, DC.
- [2]. JP 1-02, (2010), *Joint Publication 1-02, Department of Defense Dictionary of Military and Associated Terms*, Joint Staff, Washington, DC
- [3]. Lezhnev, S., (2005), *Crafting Peace: Strategies To Deal With Warlords In Collapsing States*, New York, NY, Lexington Books.
- [4]. Giustozzi, A., (2009), *Empires of Mud: wars and warlords in Afghanistan*, New York, NY, Columbia University Press.
- [5]. Hamalainen, P., (2008), *The Comanche Empire*, New Haven CT, Yale University Press.
- [6]. United States Joint Forces Command Joint Futures Group, (2010), *The Joint*

- Operating Environment (JOE)*, US Joint Forces Command, Suffolk, VA.
- [7]. Wilensky, U., (1999), *NetLogo*. <http://ccl.northwestern.edu/netlogo/>, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

3.8 Using Random Forest Models to Predict Organizational Violence

Using Random Forest Models to Predict Organizational Violence

Burton Levine, Georgiy Bobashev
Research Triangle Institute

Blevine@rti.org, Bobashev@rti.org

Abstract. We present a methodology to access the proclivity of an organization to commit violence against non-government personnel. We fitted a Random Forest model using the Minority at Risk Organizational Behavior (MAROB) dataset. The MAROB data is longitudinal; so, individual observations are not independent. We propose a modification to the standard Random Forest methodology to account for the violation of the independence assumption. We present the results of the model fit, an example of predicting violence for an organization; and finally, we present a summary of the forest in a “meta-tree.”

1.0 INTRODUCTION

We used Random Forest modeling in the context of social theories to predict the probability that an organization will commit violence against non-government security forces or civilians. By leveraging existing social theories to build predictive models our results have more face validity than an approach that is exclusively data driven. Effective implementation of these models could aid intelligence analysts in accessing terrorist threats. This paper is limited to the use of one Random Forest model. Our data source is longitudinal; consequently, implementation of the Random Forest model is complicated by the correlations within the data. We present methodology used to account for this correlation in the defining of nodes within a tree. Finally, we present the results of the model fit.

2.0 SOCIAL THEORIES

In social research, it is often unacceptable to apply modeling to data without a conceptual model. This is prudent for two reasons. First, the modeling procedures can aid in identifying associations between the outcome and causal variables but cannot identify causality. Consequently, we first identified the hypothesized causal relationships and then use the data to corroborate the conceptual model. Second, if we did not limit our analysis to the factors that we believe *a priori* are related to our outcome than we are likely to present spurious results. We limited the set of organizational characteristics we considered in our models to the characteristics that had

been identified as relating to violence in the following papers [1-4].

3.0 DATA

We conducted this analysis on The Minorities at Risk Organizational Behavior (MAROB) dataset. MAROB contains information about the characteristics of ethno-political organizations that are most likely to employ violence and terrorism in the pursuit of their perceived grievances with local, national, or international authority structures. The data set contained information on 113 groups located in the Middle East and North Africa. The MAROB dataset covers the period between 1980 and 2004. Additional information about MAROB can be found at the Minorities at Risk (MAR) Web site: <http://www.cidcm.umd.edu/mar>.

4.0 RANDOM FOREST

4.1 Overview

We used *Random Forest* methodology to create large quantities of decision trees [5]. These trees are created by identifying organizational characteristics associated with violence. Then the trees are used to estimate the probability of violence given specific combinations of organizational characteristics. Random Forest methodology provides a mechanism for creating many decision trees using the same data source and the results are averaged over the trees. The two major advantages of the random forest

methodology compared to using a single decision tree is a reduction in variability and a mechanism to validate the results.

4.2 Random Forest methodology

The Random Forest model is validated by allocating the sample into training and testing portions. We fitted a Random Forest model to the training portion. Then we quantified the model fit by comparing predictions from the Random Forest model between the training and the testing datasets. Organizations have a 75% chance of being assigned to the training dataset and 25% chance of being assigned to the testing dataset. Eighty-six of the 113 organizations were assigned to the training dataset.

Using the training dataset, 100 datasets were created by sampling 86 organizations with replacement. A decision tree was created for each of the 100 datasets. The name *Random Forest* is descriptive of the methodology since many decision trees are created by randomly sampling a dataset. Each decision tree partitions the dataset. A partition of a set was a division into non-overlapping and non-empty parts that cover the entire set. These subsets were both mutually exclusive and exhaustive of the set being partitioned. And, each partition had an associated predicted probability of violence. Consequently, for each tree, an observation on our data set (organization by year combination) fitted into one and only one leaf with an associated probability of violence based on the characteristics of the organization in that year.

4.3 Growing a tree

We limited our search for characteristics to use as nodes for *growing* each decision tree to variables that were identified in our literature review of social theories describing violence in the MAROB dataset. All of the independent variables were either binary or made into binary variables by creating indicator variables for

organizational characteristics that were multinomial.

To create a tree from one data set we implemented the following procedure. First, for each variable considered, we fitted a generalized estimating equation (GEE) using the variable in question as the independent variable and next year violence as the dependent variable. Since the outcome is binary we used a logistic

$$\text{regression model: } \log \left(\frac{E[Y_{ij}]}{1 - E[Y_{ij}]} \right) = x'_{ij}.$$

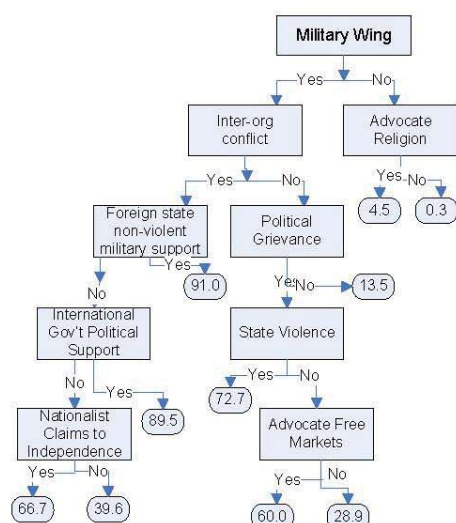
The following expression for the variance incorporates clustering: $V(\alpha) = \phi A^{1/2} R(\alpha) A^{1/2}$

where $A(\mu)$ is a diagonal matrix of individual variances and ϕ is the scale parameter. We considered two type of correlations matrixes: exchangeable and autoregressive. We chose autoregressive because we believed that correlation of data within an organization is stronger for observations that are closer in time.

We determined which of the organizational characteristics was most strongly associated with violence by picking the characteristic that contained the lowest p-value from the GEE models. This characteristic is the first node. For the second node we divided the dataset into two groups based on the characteristics of the variable in the first node and calculate the characteristic that is most strongly related to violence on that subset of the data. Repeating this procedure for nodes 3 through 5 we identified the characteristic that was most strongly associated with violence on each subset of the data. We terminated this process when there were less than 20 observations in a node of the tree.

The following is an example of one of the decision trees in the Random Forest.

Figure 1: Example of a decision tree



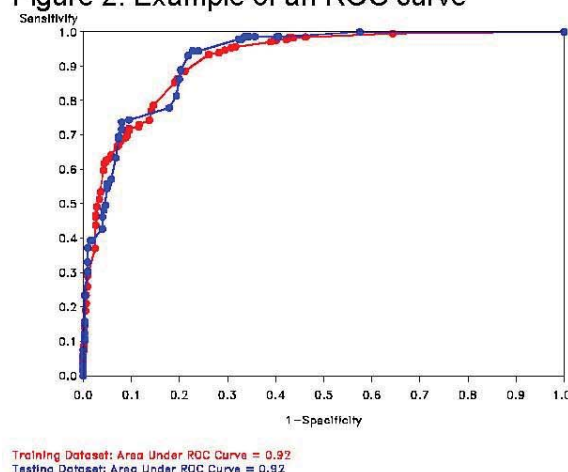
The tree above has 10 leafs. To assign a probability of violence to each leaf (the number in the ovals in the tree above) we created a variable that identified the leaf for every organization by year combination. Then we ran a GEE model with the “leaf” variable as the lone independent variable and with violence next year as the dependent variable.

The decision tree created a partition of the dataset such that each data point (each organization by year combination) was located in one and only one of the leaves of the decision tree. And, each leaf had an associated probability of violence. We used each tree to associate a probability of violence for each observation on the dataset based on the characteristics that made up the leaves of the tree. So, for each observation we generated 100 predictions, one for each tree, of the probability of violence in the next year. We then averaged over the 100 values to generate one prediction of violence. The trees were created using the training dataset exclusively. But, we generated predictions of violence for both the training and the testing dataset. Finally, we tested the quality of the predictions by generating ROC

curves for the training and testing datasets as described in [6].

Below are graphs of the ROC curves. The area under the ROC curve for the training dataset is 0.92. This indicates that the model fits the data very well. The area under the ROC curve for the testing dataset is also 0.92. This indicates that the model is applicable to data that was not used to generate the predictions. In other words, we have obtained evidence that the model is generalizable.

Figure 2: Example of an ROC curve



4.4 Meta-tree

Using Random Forest methodology to predict the probability of violence has a beneficial characteristic of being robust. By sampling a large number of datasets the potentially deleterious effect of aberrant results is minimized. However, the prediction rule is very complicated since it requires a large amount of trees, 100 in our analysis. Consequently, in its present form Random Forests are not useful for identifying the combinations of characteristics that are related to violence. But, we can use the 100 trees to create a *meta-tree* that is useful in identifying the interactions that are most common to the 100 trees. The following table displays the distribution of the first node for the 100 trees.

Node	Frequency
Military Wing	86
Inter-organizational conflict	10
Other	4

In our meta-tree we have selected Military Wing as the first node.

The following tables display the distribution of the second node for the 86 trees that contained military wing in the first node.

First node: No Military Wing	
Second node	Frequency
Foreign state non-military support	34
Dominant economic grievance	34
Other	18

First node: Military Wing	
Second node	Frequency
Inter-organizational conflict	50
State violence	16
Other	20

In our meta-tree, for the second node under no military wing, we have the same number of trees using the nodes foreign state non-military support and dominant economic grievance. Both of these nodes lead to terminal leaves. So, we considered both possibilities for our meta-tree. We selected inter-organization conflict as our second node under military wing.

The follow table displays the distribution of the third node for the 50 trees that contained “has military wing” in the first node and inter-organizational conflict in the second node.

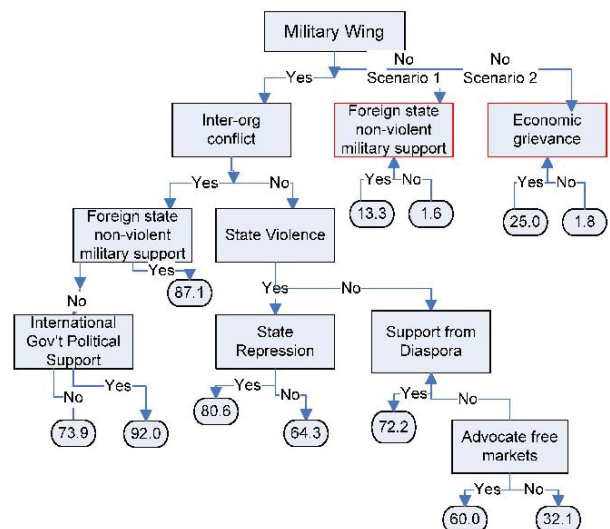
First node: Military Wing Second node: Inter-org conflict	
Node	Frequency
Foreign state non-military support	17
Received support from diaspora	13
Other	20

First node: Military Wing Second node: No inter-org conflict	
Node	Frequency
State violence	33
Other	17

We chose “foreign state non-military support” for the node under (Military wing=yes) and (Inter-organizational conflict=yes). And, chose “State violence” for the node under (Military wing=yes) and (Inter-organizational conflict=no).

Continuing the process we arrived at the following meta-tree. The boxes in red are two equally likely scenarios.

Figure 3: Example of a meta-tree



4.5 Making predictions

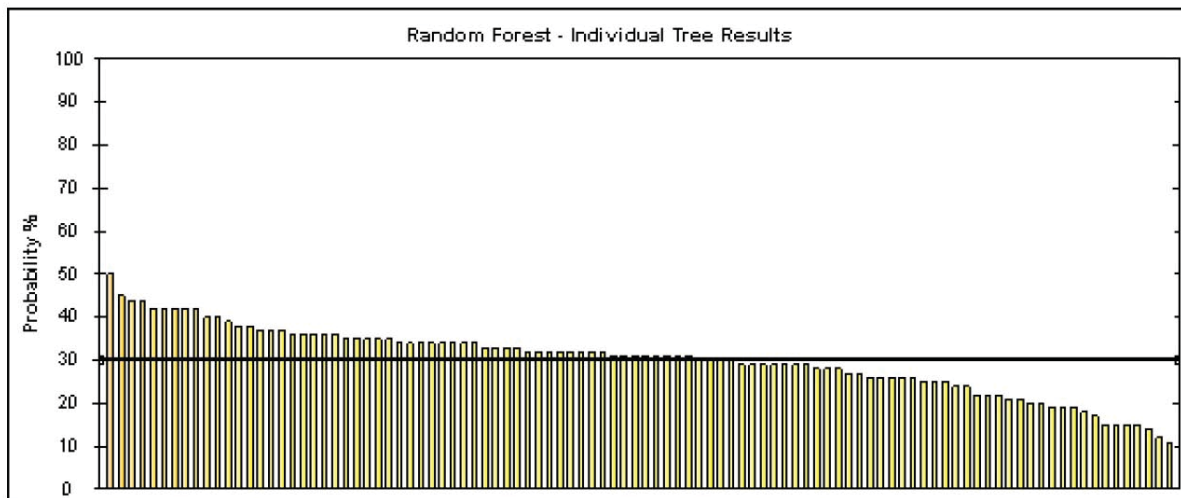
To use the Random Forest model to make predictions on an organization we need to have data on that organization. In the following table, the left column contains the nodes used in all 100 trees. The right column contains the value for the organization of interest. Rather than concoct a fictitious example, we used the MAROB data for the organization Fatah al-Intifada in 2004. We made predictions on whether this organization will commit violence on non-government personnel during 2005.

Criterion	Fatah al-Intifada 2004
Advocate free market policies	0
Dominant political grievance focused on creating separate state	1
Economic grievance focus on elimination of discrimination	0
Economic grievance of organization-remedial policies	0
Ethnic organization no claims of independence	1
Foreign state non-violent support	0
Foreign state support in current year	0
Has military wing	1
Inter organizational conflict	0
Inter organizational conflict severity	0
International government political support	0
International government organizational support	0
International non-government organizational support	0
Intra-organizational conflict	0
Organization is legal	1
Nationalist claims of independence	0
Organization advocates policies concerning religion	0
Organization open	1
Received support from diaspora	0
Repression by state	0
State violence against organization	0

To calculate this probability, for each of the 100 trees, we found the node associated with the organizational characteristics of Fatah al-Intifada in 2004 and the value of the probability of violence in that node. Then

we averaged the probabilities for such nodes over all trees. It turned out that the probability of violence in the next year was 30%. A bar graph in Figure 4 presents the results of the prediction with the Random Forest model. Each vertical bar represents the prediction from one tree and the horizontal line is the average value over all 100 trees. An analyst can also explore individual tree models by examining the groups that form trees at either sides of the graph. For example, on the left hand side there are trees that produce high probability of violence (almost 50%). On the right end of the graph the trees contain groups that have lower probabilities (around 10%). Because each prediction is based on groups that share the same characteristics of the node with Fatah al-Intifada, one might want to take a closer examination of these groups' characteristics to think of possible scenarios of violent or not violent solution of the situation.

Figure 4. Bar graph of predicted probabilities for trees in the Random Forest predictive model



5.0 Discussion

Although it is not possible to predict the future, it is often possible to assess risks and probabilities of certain events to happen. All these forecasts and predictive probabilities carry a load of uncertainty error. Most of these errors we cannot control because model is always a simplification of the reality, however Random Forest methodology allows one to assess the within-model error, i.e. assuming that the model is correct. As most forecasting techniques this model is based on similarity rationale, i.e. similar (in some sense) groups would behave in a similar way. No two groups are exactly the same and a selected group could be similar to a variety of groups in different ways e.g., size, political goals, religious views, legal status, etc. The advantage of our Random Forest approach is that it examines a broad variety of similar groups, e.g. groups that are similar in many characteristics but are either very violent or not violent at all. Focusing on these opposite sides of the spectrum allows an analyst to assess what *additional* characteristics or government actions contribute to similar groups shift towards or away from violence.

In the intelligence analyst setting, assigning precise estimates are usually avoided. As a result, when these models are implemented the probability estimates will be mapped into the following categories: high, medium and low probability of violence.

The implementation of this methodology has several limitations. The results will only be valid for organizations that will fit the MAROB inclusion criteria. These inclusion criteria are described in the MAROB website. Furthermore, one must know a lot of information about the organizational characteristics of the group one is interested in to apply this model. If the user of the model is unsure of some of the characteristics they can either take their best guess or input all combinations of the missing data. Each combination will result in a different predicted probability of violence. Using these predicted probabilities one can create a range of predicted probabilities.

6.0 References

- [1] Asal, V. (2007, August). *Organizational factors and the choice of terrorism: Minorities at Risk Organizational Behavior in the Middle East*. Paper presented at the annual meeting of the American Political Science Association, Chicago, IL.
- [2] Post, J., Ruby, K., & Shaw, E. (2002a). The radical group in context: 1. An integrated framework for the analysis of group risk for terrorism. *Studies in Conflict and Terrorism*, 25, 73–100.
- [3] Post, J., Ruby, K., & Shaw, E. (2002b). The radical group in context: 2. Identification of critical elements in the analysis of risk for terrorism by radical group type. *Studies in Conflict and Terrorism*, 25, 101–126.
- [4] McCauley, C. (2009). The 21-item threat criterion list: Review and suggestions. In M. Schwerin, *Violent Intent Modeling and Simulation (VIMS) year-end report*. Prepared for the Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security. Research Triangle Park, NC: RTI International. Appendix E.
- [5] Breiman, L. (2001). Random Forests, *Machine Learning*, 45, 5-32.
<http://www.springerlink.com/content/u0p06167n6173512/fulltext.pdf>.
- [6] Zweig, M.H., Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, 39: 561-577.

7.0 Acknowledgements

This work was supported by the VIMS contract with the Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security.

The corresponding author is Burton Levine, (919) 541-1252; E-mail: blevine@rti.org; Contract Number HSHQDC-08-C-00100. The views expressed in this article are our own and do not reflect the official policy or position of the Department of Homeland Security or the U.S. Government. The study is approved for public release for unlimited distribution.

3.9 A Model of Rapid Radicalization Behavior Using Agent-Based Modeling and Quorum Sensing

A Model of Rapid Radicalization Behavior Using Agent-Based Modeling and Quorum Sensing

Noah Schwartz, Ph.D.
Christopher Newport University
noah.schwartz@cnu.edu

Nick Drucker
MYMIC, LLC.
nick.drucker@mymic.net

Kenyth Campbell
Old Dominion University
kenyth.campbell@gmail.com

Abstract: Understanding the dynamics of radicalization, especially rapid radicalization, has become increasingly important to US policy in the past several years. Traditionally, radicalization is considered a slow process, but recent social and political events demonstrate that the process can occur quickly. Examining this rapid process, in real time, is impossible. However, recreating an event using modeling and simulation (M&S) allows researchers to study some of the complex dynamics associated with rapid radicalization. We propose to adapt the biological mechanism of quorum sensing as a tool to explore, or possibly explain, rapid radicalization. Due to the complex nature of quorum sensing, M&S allows us to examine events that we could not otherwise examine in real time. For this study, we employ Agent Based Modeling (ABM), an M&S paradigm suited to modeling group behavior. The result of this study was the successful creation of rapid radicalization using quorum sensing. The Battle of Mogadishu was the inspiration for this model and provided the testing conditions used to explore quorum sensing and the ideas behind rapid radicalization. The final product has wider applicability however, using quorum sensing as a possible tool for examining other catalytic rapid radicalization events.

1. INTRODUCTION

What is the cause of rapid social and political radicalization? Over the past several years, rapid radicalization has caught the attention of governments because of the potency of radicalized movements and the increased prevalence of radicalized groups. Although studying these types of events after the fact may help shed light on the dynamics of the events, it does not provide much insight into the underlying factors that cause radicalization. Because rapid radicalization events occur so quickly, often with no warning at all, and usually dissipate equally as quickly, within anywhere from a couple of hours to a couple of days, it is very difficult to collect usable data from these events. However,

recreating events using modeling and simulation (M&S) allows researchers to study some of the complex dynamics associated with rapid radicalization.

Agent Based Modeling (ABM) is an M&S paradigm well suited for modeling group behavior and for simulating factors that can lead to rapid radicalization. In this paper, we introduce the concept of quorum sensing within an ABM framework as a means of exploring rapid radicalization and as a possible explanation for why and how rapid radicalization can occur.

In biological systems, quorum sensing refers to a decentralized decision making process in which the behavior of a group is correlated with the population density of that group [5]. During a quorum sensing decision, each individual within the group emits and receives a signal from other group members; the behavior of each

individual changes when the accumulated signal exceeds a threshold. Due to the reciprocal nature of quorum communication, group behavior often changes rapidly, resembling the rate of change seen in human social and political radicalization.

In the current paper, we will extend a basic quorum sensing model to the domain of human interaction in a volatile social environment. We hypothesize that quorum sensing will result in more instances of rapid radicalization, and the component parts of quorum sensing will help describe dynamic group behavior under stressful conditions.

2. BACKGROUND

For the purpose of this paper, we define rapid radicalization as a process in which a small subset of a populace join to perform some radical action; this process spreads quickly until a majority of the population is participating. This term is similar to an event where military personnel will move synchronously to a target or point of contact. Unlike coordinated military movements, however, rapid radicalization occurs despite the presence of a central organizing figure, no defined lines of communication between mobilizing individuals, and no clear process for deciding when individuals should move or take action. Despite these differences, the outcome of organized military movements and rapid radicalization can be very similar.

An example of rapid radicalization can be seen in the Battle of Mogadishu, more famously known as Black Hawk Down. During the battle, US forces were primarily engaged with hostile forces loyal to local warlords. When the battle first began, these factions organized quickly and converged on the location of two downed US Blackhawk helicopters[1]. Over the course of the battle, however, many non-militant forces mobilized in a similarly fast and coordinated manner, taking up arms against the US Soldiers. These non-militant forces included members of small independent groups that were established to protect city blocks from the warlord forces. Within a

matter of hours, these independent groups were mounting an apparently unified attack on US Soldiers.

Rapid radicalization events such as this raise questions about the motivation and communication that allow independent groups to operate in a unified, coherent manner. One anecdotal story from the book "Black Hawk Down" by Mark Bowden recounts how a young man was caught in the crossfire between US forces and warlord forces. The young man was a member of the civilian population and was shot and killed by US Soldiers. A friend of the boy witnessed his death and immediately returned home to retrieve his rifle, meant to defend himself from the warlord groups, and returned to engage in the firefight against the US Soldiers. This individual reached a tipping point at which time he felt compelled to act in a radical manner.

Another example of rapid radicalization can be seen in the recent wave of revolutionary uprisings in the Middle East and North Africa, known as the "Arab Spring." Since December 2010, large-scale demonstrations and civil resistance movements have occurred in over 16 countries [2]. Although these demonstrations have been independent, targeting specific issues in each local country, the movement as a whole appeared to share a common momentum and to operate with a unified purpose. Social networking sites are credited with providing a relatively stable and secure communication platform to facilitate the organization of these movements and to broadcast news of the events as they occurred [7]. However, access to these sites alone does not explain the speed and magnitude of self-organization within each individually motivated population acting against similarly independent governments.

We propose that in rapid radicalization situations, there are underlying factors that affect each individual; communication media such as social networking sites simply provide a tool each individual can use to observe the state of others in a similar

situation. When a person is motivated to act against a larger force to change their situation, taking action may expose that person to an unacceptable level of risk; when a quorum arises, the person no longer faces this risk individually and enjoys “safety in numbers.” We believe that this type of decision process characterizes rapid radicalization. The goal of this paper is to explore the factors that contribute to this decision process and to understand what compels an individual to commit to a radical course of action that, when taken as an individual, is prohibitively dangerous.

2.1. Quorum Sensing

In biological systems, quorum sensing refers to a decentralized decision making process in which the behavior of a group is correlated with the population density of that group [5]. Quorum sensing operates through a reciprocal communication process in which each individual in the population signals its presence to its neighbors, and each individual detects the presence of its own neighbors. The detected signal is therefore directly proportional to the population density. When the quorum population density exceeds a threshold, behavior of each individual in the population changes very rapidly. For this reason, we have chosen to use a mathematical model of quorum sensing as the basis for the current model of rapid radicalization.

When examining human behavior as a quorum-sensing decision within a population, one must make several assumptions about each individual in the population. First, in order for quorum sensing decisions to occur, each individual in a population must have some capacity for signaling their presence and for detecting the presence of other elements. Second, quorum sensing assumes that each individual in a population is motivated to perform some action, but taking action under the current conditions is not advantageous for the individual due to a high cost of action and perhaps a high likelihood of failure or economic loss.

The current paper evaluates a quorum-sensing model that is based on passive, spatial and temporal diffusion of information. While the focus of the current paper is rapid radicalization, results will focus on the time of first radicalization, the duration of radicalization from first attack, and the total time for full population radicalization to occur from the initial state of the model.

2.2. Line-of-Sight

This paper also includes a default radicalization model based on line-of-sight contact with radicalization events such as death or loss of/damage to valued assets. The motivation behind this model was the radicalization of neutral civilian forces observed in the Battle of Mogadishu, described earlier. In this model, individuals in the population decide to attack in response to a death event occurring within an individual's line of sight. Following initial mobilization, the individual will influence other neighbors to join the attack.

This model differs from the quorum-sensing model in several respects, specifically, the style and speed of communication between individuals, the motivation of individuals in the simulation, and the catalyzing events that lead to radicalization. Unlike the quorum-sensing model, the line-of-sight model assumes that, by default, individuals are not motivated to attack. Individuals will only take action and join the attack after witnessing a catalyzing event such as a death or by coming into close contact with a neighbor who is fighting. This assumption is made deliberately so that the model will be devoid of past influences, i.e. the majority of the population is politically neutral and has no preexisting bias toward action or inaction without the target posing a real threat. The quorum-sensing model, on the other hand, assumes that individuals are motivated to attack, but choose to remain passive because of the prohibitively high personal risk that is inherent in attack.

3. METHODS

All models were created using an ABM framework with the NetLogo platform [6], an ABM environment developed in the Java language. In ABM, each individual is represented as an agent, operating independently and in parallel with other agents in the model. The framework can also include real world constraints and life events [3]. Agents in these models behave according to a set of rules specified by the programmer and then interact with one another according to this rule set, recreating complex human interactions [4]. Researchers often employ ABM because it allows them to control precise individual behaviors and to monitor the aggregate behavior of a group for the purpose of experimentation and study. We chose to use this framework because we are interested in the activity of multiple individuals within a group. ABM allows us to create this group and to provide discrete rule sets describing quorum sensing and line-of-sight behavior.

We selected an M&S approach because it is not feasible to examine rapid radicalization as it unfolds, interviewing individuals and collecting data. M&S provides a means to simulate these events in a highly controlled and replicable fashion, repeating simulations and modifying details of the simulations on subsequent runs.

3.1. Model Construction and Behavior Specification

Both models consisted of 300 agents distributed randomly across a 33x33 grid of "patches." During each simulation, agents interacted on this grid over up to 24 time steps, with each time step arbitrarily defined as equivalent to 1 hour. During each time step, agents cycled through four behaviors: broadcast status, sense status, fight, and move.

Each agent is described by several variables: *attack status*, *agitation level*, *attack threshold*, and *spatial position*. By default, all agents begin in a non-attack

status and with agitation level equal to zero. Agitation level ranges from 0 to 1, and attack threshold also ranges from 0 to 1. By default, the attack threshold is equal to 0.9 +/- 0.1. Agents in the quorum-sensing model contained an additional *risk* parameter defining the percent-level risk associated with attack (0% = no risk, 100% = maximum risk).

At the beginning of each time step, agents broadcast their agitation level to other agents; the broadcast method varies between quorum-sensing and line-of-sight models and is described in more detail below. During the sense stage, agents detect environmental events including death events as well as the agitation level of neighboring agents. These inputs increase or decrease the agents agitation level during the fight stage; if the agitation level exceeds the attack threshold, the agent goes into an attack status. At the end of each behavior loop, agents move one patch in a random direction.

Specific operations carried out during the broadcasting, sensing, and fighting stages differed between quorum-sensing and line-of-sight models. Behaviors and associated equations are described below for each model.

Quorum-sensing

broadcast: each agent emits a persistent agitation signal defined as *emission strength*, ranging from 0-1 that decays exponentially with distance and time

sense: incoming agitation signals are added; total incoming agitation is subtracted from risk and transformed into a local agitation level according to the sigmoidal decision function shown below in eq. (1) and Fig. 1.

Eq. (1)

$$local\ agitation = \frac{1}{1 + e^{(risk - incoming\ agitation)}}$$

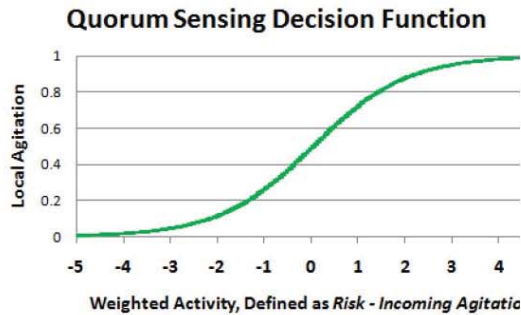


Figure 1: Sigmoidal Sensing Function Graph

fight: if local agitation exceeds threshold, switch to attack mode

Line-of-sight

broadcast: if attacking, agents communicate their attacking status to other agents within a 6-patch radius; signal is weighted according to a *social influence* parameter that ranges from 0-1 (i.e. recruiting neighbors to attack); signal decays linearly with distance

sense: incoming agitation signals are added to calculate a local agitation level

fight: if local agitation exceeds threshold, switch to attack mode; if agent is located within 8 patches of the center of the simulation and a death has occurred, set movement to false; with 20% probability, increase agitation level of link neighbors by a parameter that ranges from 0-1 (i.e. death of a family member or friend); remove the dead agent from the simulation

3.2. Experimental Protocol

For each model, we are concerned primarily with the time of first radicalization, the duration of the radicalization process, and the total time for full radicalization to occur from the initial state. These variables were measured at the conclusion of the 24 time steps, or when percentage of radicalized agents reached 100%. Parameters of each model were varied systematically in an effort to understand the role of each factor in the radicalization

process. The specific manipulations that were made were determined through a piloting process intended to discover values of each variable that would produce a complete range of the outcome variables.

In the quorum-sensing model, level of risk and emission strength were varied independently. Level of risk was set to 26.7%, 30.0%, and 33.3%; emission strength was set to 0.5 – 0.7 in steps of 0.05. These manipulations represent a change in the risk associated with action, and a change in the level of agitation that each agent expresses.

In the line-of-sight model, social influence and death influence were varied independently. Social influence was set to 0.3, 0.4, 0.5, and 0.6; death influence was set to 0.5, 0.6, 0.7, and 0.8. These manipulations represent a change in the affect an attacking agent has on neighboring agents, and the affect a death event has on nearby agents.

To account for random outliers in simulation results, each simulation was run 10 times for each manipulation and the results below consist of the average outcome across all 10 iterations.

4. RESULTS

The quorum-sensing and line-of-sight models were explored parametrically in order to evaluate the radicalization effects produced by each. Rapid radicalization is defined as a process in which a small subset of a populace join to perform some radical action; this process spreads quickly until a majority of the population is participating. Based on this definition, the model behaviors of interest to the current paper are the time of first radicalization or attack, the net duration of radicalization process, and the total time for full population radicalization to occur from the initial state. These data are shown graphically in figs. 2-7 below.

4.1. Quorum-Sensing

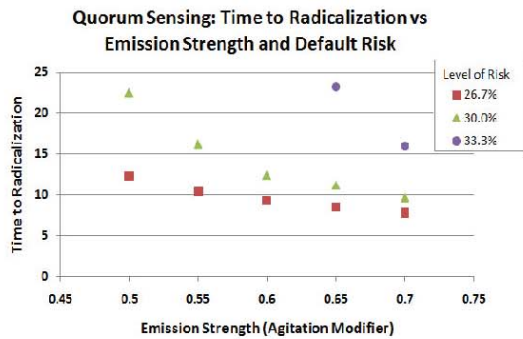


Figure 2: Total time to radicalization in the quorum-sensing model, from the initial state of the model for varying levels of risk and emission strength (magnitude of agitation expressed).

Figure 2 shows that at 33.3% risk, emission strength below 0.65 did not result in rapid radicalization suggesting that the desire to express feelings of agitation was outweighed by the risk associated with taking action.

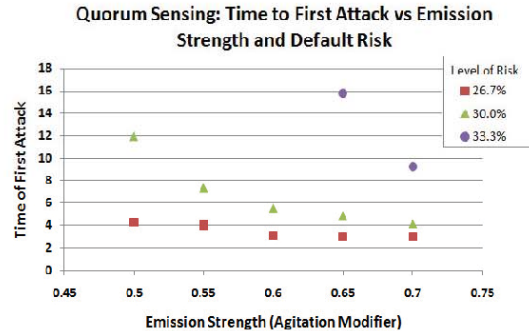


Figure 3: Time to first attack in the quorum-sensing model.

The numbers in Fig. 3 reflect the speed with which the population begins to become radicalized. For example, during the Battle of Mogadishu, the population became radicalized within 4-6 hours after the battle began suggesting that individuals had little to lose by joining the battle.

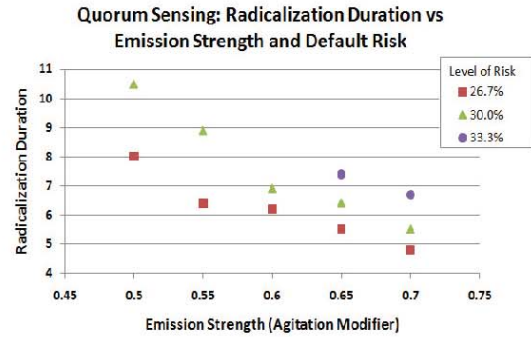


Figure 4: Radicalization duration from the time of first attack in the quorum-sensing model.

Figure 4 reflects the rate of radicalization in the population: the lower the number, the faster the population (N = 300) became fully radicalized. These results suggest that the population in the Battle of Mogadishu was highly expressive in their agitation.

4.2. Line-of-Sight

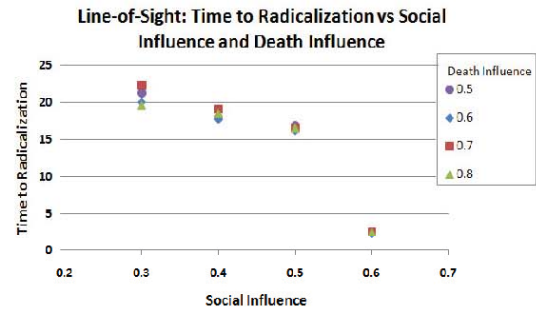


Figure 5: Total time to radicalization in the line-of-sight model, from the initial state of the model for varying levels of social influence and death influence.

Figure 5 demonstrates time to radicalization varies as a function of social influence, but is unaffected by changes in death influence.

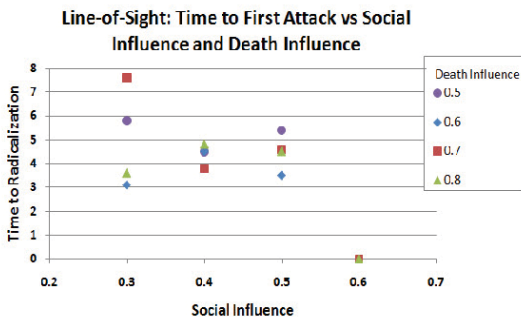


Figure 6: Time to first attack in the line-of-sight model.

Figure 6 demonstrates that similar to the total radicalization time, high social influence seems to lead to rapid radicalization whereas there is little affect of death influence. In the data shown above, the effects of death influence are unstable when social influence drops below 0.6

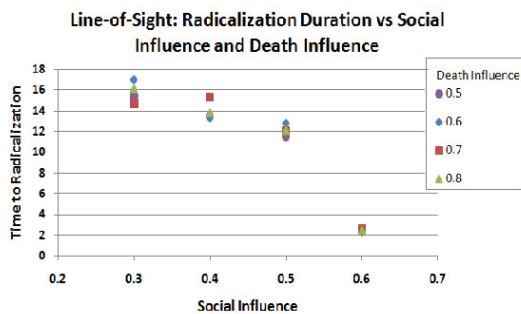


Figure 7: Radicalization duration from the time of first attack in the line-of-sight model.

The rate of radicalization in Fig. 7 appears to depend primarily on social influence, not on death influence. This is not surprising given that, in the current simulation, death was the catalyzing event and only affected those agents nearby, whereas social influence played a continuous role in the radicalization of other members of the populace.

5. CONCLUSION

The goal of the current paper was to explore the radicalization of a population as modeled using quorum sensing and line-of-sight influence. As shown in the figures 1-3 above, changes in both risk and expressed agitation yielded visible changes in the

onset and speed of rapid radicalization. These changes decrease monotonically with both increasing risk and increasing agitation. As shown in figures 4-6 above, line-of-sight influence varied significantly with changes in social influence, but there were no visible changes apparent with change in death influence. The only exception to this is instability in time to first attack during situations where social influence is weak (fig. 5). It is possible that this instability occurred because line-of-sight influence is highly sensitive to the spatial distribution of agents; if a death occurred in a relatively sparse area, the time to first attack would be substantially increased.

Given the dynamics demonstrated in the current model, quorum sensing or line-of-sight (or some combination of the two) may be used as a basis for modeling other rapid radicalization events such as violent insurgency, political uprising, and other forms of radical action. Such movements have occurred numerous times, both recently and in more distant past. Recent insurgency may provide a model for understanding radicalization as it occurs in spatially-diffuse networks with unreliable communication. Similarly, older examples such as the revolutions of 1989 in Poland and Eastern Europe and even the revolutions of the 18th and 19th centuries (notably, the US and France) may provide examples of slow radicalization occurring over extended distances and timescales.

Using a common modeling framework for each of these older cases would allow one to draw interesting comparisons between older revolutionary periods and the current Arab Spring, especially in examining the role that modern social networking sites have played in facilitating communication between individuals. The results of the experiment further confirm this as the line-of-sight model demonstrated high sensitivity most likely owing to the distance between agents. The socially connected nature of many members of the Arab states experiencing uprisings would remove the limitations imposed on agents in the line-of-sight model.

These models may also be augmented to include networks of unique networks such as independent countries seeming to act in parallel, or with competing populations who vary in terms of their decision priorities (for example, honor-based versus reputation-based societies). Finally, it may also be possible to adapt the models to situations involving fewer agents in a spatially enclosed area, such as a prison riot.

While the performance of these models suggests the promise of explanatory power, these models can also be used to understand how one may disrupt and/or facilitate the radicalization of the population through methods such as communication disruption, quorum quenching, or systematic agitation; manipulations that are possible to explore within an ABM/M&S framework.

6. REFERENCES

- [1]. Bowden, M. (1999). *Black Hawk Down: A Story of Modern War*. Grove Press: New York.
- [2]. Doran, M. S., (2011), *The Heirs of Nasser*, *Foreign Affairs*, Vol 90 (3), p17-25.
- [3]. Epstein, J. (2006). *Generative Social Science Studies in Agent-Based Computational Modeling*. Princeton University Press: Princeton, New Jersey.
- [4]. Miller, J. H., & E. Page, S. E. (2007). *Complex Adaptive Systems: An introduction to Computational Models of Social Life*. Princeton University Press: Princeton, New Jersey.
- [5]. Miller, M., & Bassler, B. (2001). Quorum sensing in bacteria. *Annual Review of Microbiology*, 55, 165–99.
- [6]. Wilensky, U. (1999). *NetLogo*. <http://ccl.northwestern.edu/netlogo/>, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
- [7]. Williamson, A. (2011), *Social Media and the New Arab Spring*, Hansard Society, retrieved July 13, 2011 from <http://www.hansardsociety.org.uk/blogs/edemocracy/archive/2011/04/19/social-media-and-the-new-arab-spring.aspx>

3.10 The Human Dimension of Closing the Training Gap for Fifth-Generation Fighters

The Human Dimension of Closing the Training Gap for Fifth-Generation Fighters

Jaclyn Hoke, Alex Postnikov
Rockwell Collins, Inc.

jahoke@rockwellcollins.com
apostnikov@rockwellcollins.com

Thomas Schnell, Ph.D.
Operator Performance Laboratory (OPL), University of Iowa
thomas-schnell@uiowa.edu

Abstract. Based on a review of the recent technical literature there is little question that a serious training gap exists for fifth-generation fighters, primarily arising from the need to provide their own red-air. There are several methods for reducing this gap, including injecting virtual and constructive threats into the live cockpit. This live-virtual-constructive (LVC) training approach provides a cost effective means for addressing training needs but faces several challenges. Technical challenges include data links and information assurance. A more serious challenge may be the human factors dimension of representing virtual and constructive entities in the cockpit while ensuring safety-of-flight. This also needs to happen without increasing pilot workload. This paper discusses the methods Rockwell Collins and the University of Iowa's Operator Performance Lab use to assess pilot workload and training fidelity measures in an LVC training environment and the research we are conducting in safety-of-flight requirements of integrated LVC symbology.

1.0 INTRODUCTION

In early 2011, RAND Corporation published a Project Air Force report titled "Investment Strategies for Improving Fifth Generation Fighter Training" that analyzed the current state of training for fifth generation fighters. Using a variety of data sources the authors concluded that a significant training gap exists. The report also stated that bringing training to acceptable levels would require significant investment by the services.

One of the major factors indicating a training gap exists is the so-called "Insatiable Demand for Red Air", or opposing forces. F-22 units are required to provide their own red-air component, meaning the few v. many training exercises utilize almost all available aircraft. To further complicate the matter, using similar aircraft as red air degrades the training received by blue air pilots. Other factors contributing to this conclusion are range and airspace limitations that result from increasing airspace requirements for commercial and general air traffic and the lack of documented training requirements. It is likely that these factors will only increase as the F-35 enters the operational inventory.

These conclusions lead to the obvious question, "How do we close the training gap?" Five options were discussed in the RAND 2011 report, and analysis of the costs associated with each option led the authors to conclude "in the long run, development of the Live-Virtual-Constructive (LVC) ability to inject simulated and constructive threats into live aircraft may be the only fiscally responsible approach to improving training" [1]. Calculating the return on investment for developing this capability is complicated. The technology to accomplish this is not yet fully developed, and as such, the actual costs associated with it are unknown.

Cutting funding for live training to enable significant investments in LVC is not the answer. In fact, this would only widen the training gap. Instead, smaller, strategic investments must be made to address the challenges and mature the technology before any significant investment is made.

2.0 HUMAN FACTORS OF LVC

In order to illustrate our perception of what LVC is, the challenges we associate with it, and our approach when performing research in this domain, we propose the taxonomy shown in Figure 1. We have

proposed this taxonomy with the hope that others will add to or challenge its applicability.

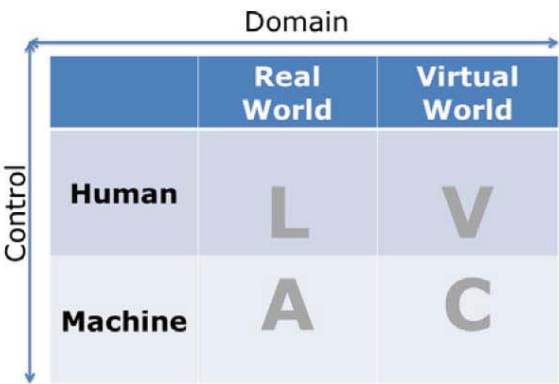


Figure 1. Live, Virtual, Constructive and Autonomous as related to the worlds and executors of actions in those worlds

The taxonomy illustrates that there are humans and computers executing actions in both the real and virtual worlds. As true LVCA (A represents autonomous systems) environments become a reality, the line between the real and virtual worlds will be blurred, requiring humans and computers to interact in a new, blended environment. Operating in this blended environment gives rise to safety challenges described further in the following sections.

2.1 Challenges

Whereas LVC can provide a cost effective means for addressing training needs, there are several challenges that must be overcome. Researchers active in this domain are quite familiar with two of the key technical challenges, data links and information assurance, but the impact on the operators has yet to be studied in sufficient detail. We will touch briefly on each of these technological challenges, and how they relate to the human dimension, and then discuss the human factors in more depth.

2.1.1 It Is (Almost) All About Connectivity

As the section title states, one of the biggest obstacles is the ability to connect

participants. The complete end-to-end connection of participants is at the crux of cracking the LVC challenge. If there are no meaningful ways to enable information exchange between the players, then discussion of the LVC topic is moot. The virtual and constructive threat information cannot be injected into the live platform if it cannot reach it.

Unlike virtual platforms on the ground that can be easily connected via optical fiber, the live platform data links must go over the radio. The real world presents challenges for maintaining connectivity at long distances when the live participants move at various speeds, with different orientations with respect to receiving antennas, etc. The sheer volume of data required for real-world scenarios also presents a significant obstacle and drives the requirements for data link robustness, as well as traffic shaping and scheduling algorithms. It may not be appropriate or practical to burden tactical networks with training data, as is done in the approach discussed by Lechner and Wokurka [2]. Appended LVC systems may require a separate data link, allowing preservation of tactical network bandwidth for operational communications rather than for training data.

Incomplete or inconsistent representation of virtual/constructive training data may be the result of data link latency, loss of data link, or data routing issues. Large entity counts that may be generated by LVC federations can quickly exceed the physical limits of the data link, giving rise to the potential for lost or non-transmitted tracks. Entities that are conveyed through a data link may generate tracks that can become stale or desynchronized when the data link breaks during aircraft maneuvering or due to flight outside the data link operational envelope. Tracks that become stale or disappear due to data link outages will cause confusion in LVC participants, especially in exercises with very large entity counts. The confusion of tracks among LVC exercise participants may result in degraded training and may

also cause unsafe merges between live aircraft and non-participating traffic or interlopers.

2.1.2 If Data Can Reach the Destination, Should It?

Assuming the data link challenge was solved and traffic could flow from one participant to another, it does not mean it is *allowed* to flow, according to classification rules. This cross-domain challenge is related to the fact that the virtual world must address security concerns often present in the real world scenarios. For example, an LVC system connecting a U.S. Marine Joint Terminal Attack Controller (JTAC) at the unclassified Camp Pendleton simulation facility to the Nevada Test and Training Range, which operates a number of Secret and Top Secret aircraft platforms, would be completely useless without a process for handling the hierarchies of data distribution and crossing classification domains.



Figure 2. Cross domain solutions are a necessary component of network-centric environments, such as LVC systems, that require information sharing across all security enclaves

Because High Level Architecture (HLA), the protocol used for many distributed simulations, is based on a publish-subscribe mechanism, any participant can subscribe to information from other participants. From a security perspective, this reduces the amount of control over the data flowing between the participants and presents the risk that sensitive information about planned missions, performance capabilities (platform, sensors, weapons, etc.), locations of facilities, task force composition, tactics,

doctrines, etc. may be disclosed. Möller *et al.* [3] explore several different techniques for handling the cross domain challenge, most of which involve limiting and/or obfuscating the data being passed between the participants.

What impact does this limited, degraded, or complete lack of information have on the trainees? Is the information good enough to still provide effective transfer of knowledge? Or, perhaps more concerning, does it result in *negative* training? We pose these questions knowing that there currently are no definitive answers.

2.1.3 The Human Dimension

Whereas the integration of virtual and constructive elements into live training opens up new avenues for training, it also raises concerns about safety of flight. When needed, aircrew integrated into this blended environment must be able to differentiate between actual and artificial entities or threats. Additionally, aircrew manning live aircraft will be expected to continue to execute their original duties while coping with the changes introduced by these modifications.

The first, and probably most obvious, concern is an increase in cognitive workload - the complexity and confusion that could occur when additional information is introduced into an already complex systems management problem. Minor increases in pilot workload can have an impact on safety critical tasks, such as visual search strategy and threat deconfliction methodology. The training benefits of LVC could be easily offset if this flight safety hazard is not mitigated appropriately.

A second area of concern lies in the misinterpretation of virtual and/or constructive symbology or annunciations. Lacking appropriate safeguards, a pilot could misinterpret a live entity as a virtual input and, considering it to be a training target, perform inappropriately threatening maneuvers or obtain radar lock against a

live aircraft not participating in the training exercise. These types of actions could result in a real-world rules-of-engagement violation. Conversely, misinterpreting a virtual entity could lead to risks such as fuel emergencies as a result of flying to a simulated tanker, performing high risk maneuvers to engage or evade a virtual entity perceived to be a genuine threat, or disrupting the exercise due to a traffic conflict that does not exist in the real world.

Until quite recently, these concerns caused decision makers to be reluctant to add LVC capabilities to operational and readiness training. In certain training environments, instructors can provide a margin of safety for an overloaded student, but when an LVC environment is applied to operational training and readiness there is no dedicated instructor. Mitigating the safety of flight concerns can possibly be achieved through the use of distinct symbology and/or annunciations that allow the pilot to quickly distinguish between information based on real-world data and that generated by virtual and constructive participants. However, any additional cognitive workload induced by blended symbologies, annunciations, and training scenarios must be handled in a manner that does not further compromise pilot safety.

2.2 Quantifying the Human Dimension

Until quite recently, the human dimension has been largely overlooked. Interest in quantifying the impacts on operators has increased, as evidenced by the Office of Naval Research (ONR) Broad Agency Announcement (BAA) 11-005 [4], initially posted on December 3rd, 2010. The BAA contains three technical areas that focus on the training fidelity concerns in the live, virtual, and constructive domains respectively.

Our research team has a four year history of conducting LVC exercises integrated in the U.S. airspace, and has developed a real-time workload measurement system for use

in both aircraft and flight simulators. Although this neuroergonomic operator monitoring and evaluation system [5] [6] was developed independent from the LVC research conducted by the team, it is well suited for pilot performance assessment in integrated LVC environments. Our research team has a number of experimentation systems already in place, including two L-29 jet aircraft and several reconfigurable simulators, which are equipped with these tools to evaluate pilot workload and training effectiveness. Successful tests have already been conducted in flight simulators, research vehicles, and flight test aircraft [7].

The Cognitive Avionics Tool Set (CATS), shown in Figure 3, uses neurocognitive and physiological signals to generate a real-time measure of cognitive workload. CATS uses data from respiration, electroencephalogram (EEG), electrocardiogram (ECG), eye tracking, and galvanic skin response (GSR) sensors to measure a wide range of human neural and physiological characteristics. State-of-the art active shielding electrodes, combined with filtering and processing techniques, are used to reduce the effects of adverse noise and signal acquisition. The CATS framework was designed to easily accommodate new sensors and analysis techniques.

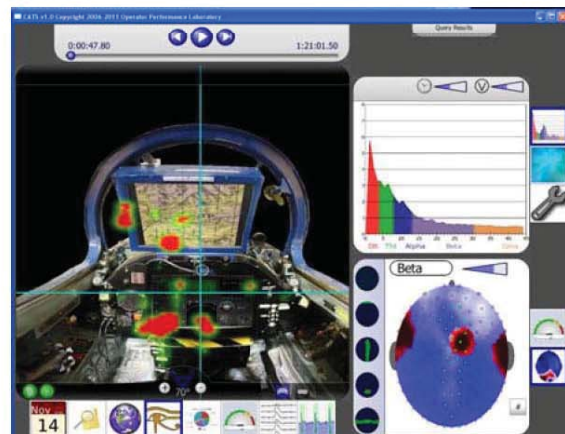


Figure 3. CATS workload monitoring system

A highly integrated and ruggedized instrumentation package, with a single point umbilical connection to the aircraft or flight simulator, simplifies the deployment of neurocognitive and physiological sensors. EEG electrodes have been integrated into the liner of a flight helmet and the respiration belt and ECG electrodes are worn under the flight suit. All peripheral electronics have been integrated into a pilot survival vest.

The Quality of Training Effectiveness Assessment (QTEA) tool [6] uses data from CATS and quantifies pilot performance against set performance requirements. Both systems record a rich dataset, containing hundreds of variables that constitute some form of measurement of effectiveness, for further offline analysis. CATS and QTEA provide powerful, state of the art analysis capabilities, and our current research plan involves applying these tools in an LVC environment to assess symbologies for injecting virtual and constructive entities into live platforms.

3.0 DISCUSSION

Workload and performance are fundamental considerations when designing new aircraft or weapons systems. To successfully augment real world inputs with training inputs coming from simulation systems the original design factors must be modified, and the viability of these modifications must be tested in the cockpit. Using both in-flight and ground-based evaluation systems, our team is conducting research into LVC-aware avionics symbology and annunciation sets and guidelines for LVC-aware avionics.

Candidate symbology design and development of simulation based prototype displays is being driven by requirements definition and stakeholder needs. Human-in-the-loop testing will gather data on the impact to safety of flight and training effectiveness of the symbologies. Tests will be conducted in an environment drawn from an application of mission tasks to a notional

LVC training scenario, and performance data will be gathered by CATS and QTEA. The tools will be deployed within our existing LVC framework to obtain real-time measures of workload, situation awareness, and pilot performance on the basis of neurocognitive and physiological measures as well as mission and flight technical performance measures. The combined system of human performance assessment tools and LVC capabilities available to our team provides a unique evaluation platform for collecting quantitative data on the suitability, workload, and cognitive demands of the proposed symbologies and annunciations.

Our experimentation system, shown in Figure 4, includes two L-29 jet aircraft with modified evaluation cockpits in the rear seat, integrated instrumentation pods, a ground support infrastructure, a reconfigurable single-seat simulator and an operator monitoring and evaluation system. This constellation of assets will allow our team to test the performance of multiple crews in an LVC exercise and conduct human-in-the-loop testing in both the reconfigurable flight simulators and LVC-enabled live aircraft. The system will also allow evaluation of the display prototypes to make strong claims about the validity and effectiveness of various virtual and constructive symbology insertion methods in a much more effective manner than traditional human factors data collection techniques, such as pilot surveys.

At this stage in our research, we foresee symbology variations using multiple stimulus dimensions for differentiating between live, virtual, and constructive entities. Variations include, but may not be limited to, symbol color, size, fill, shape, text legend, style (dashed, dotted, solid), blinking (speed, duty cycle), fading, and contrast polarity. Variations in auditory annunciations to distinguish between entity types will also be investigated, including dimensions such as voice (male vs. female), amplitude, frequency, waveform, envelope

(attack, sustain, release, decay), and stereo vs. mono. Additional variables that will be

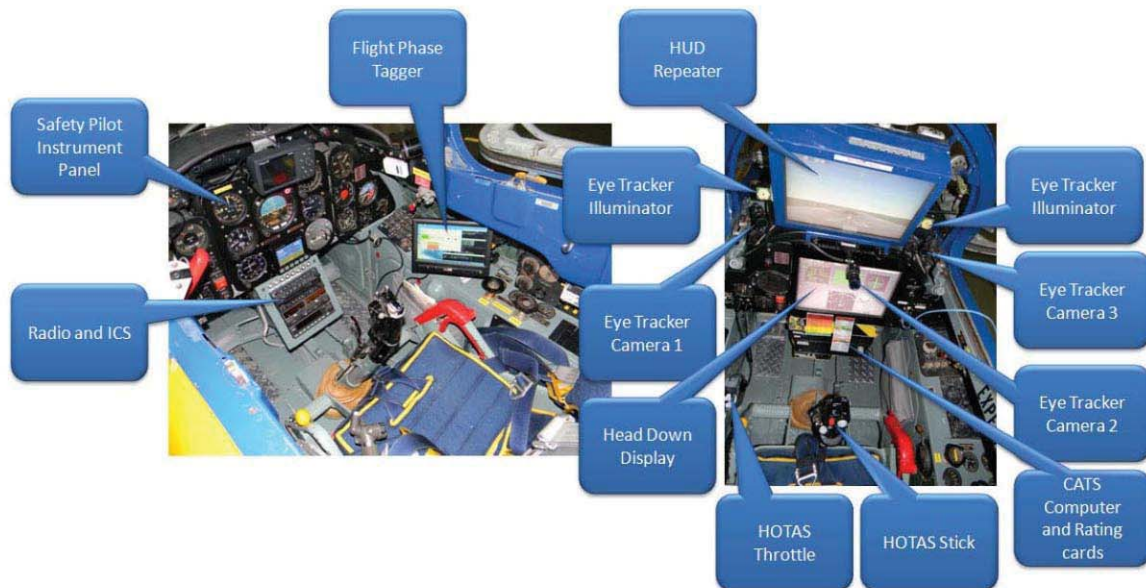
considered are type of display (visual, auditory), display location (head-up, head-



a. Pilot with Sensor Vest



b. Flight Simulator



c. L-29 Front Cockpit

d. L-29 Rear Cockpit

Figure 4. Instrumented Flight Simulator and Matching L-29 Flight Test Aircraft

down, helmet), and scenario related settings. We are also considering the potential negative impacts of allowing trainees to distinguish between live, virtual, and constructive entities, and the possibility that the highest quality of training may occur when they cannot.

As a byproduct of the design and testing efforts we plan to deliver LVC training safety

and training effectiveness assessment methods and preliminary design guidelines for LVC datalink technology. Where practical we will develop detectable criteria to indicate a potential loss of real-world situation awareness or other safety of flight concerns, and propose display modes and/or annunciation to prevent the development of flight risks that may arise when virtual and constructive entities are

injected into live systems. Our goal is to provide traceability from measured pilot performance, through analysis and design, back to the mission tasks and requirements definition. We will show how specific symbologies and avionics configurations map to the ability to safely conduct LVC training to meet training and readiness requirements. Upon completion of this research we will demonstrate that the final design meets the requirements for mission training, and has a known, negligible impact on safety of flight.

4.0 CONCLUSION

A significant training gap exists for fifth-generation fighters, and a recent study identified the LVC ability to inject virtual and constructive entities into the live aircraft as the most cost effective solution. Although the technical challenges associated with integrating live equipment into LVC environments has received significant attention, the effect on the operators has not been studied in detail. This human dimension can have a significant impact on the training benefit and acceptance of LVC for operational and readiness training.

Our research team is conducting design and experimentation work to determine LVC-specific guidelines and symbologies that enable safe and effective insertion of virtual and constructive entities into live systems. An iterative research cycle that draws upon the team's extensive human factors experience will continue to evaluate and improve symbology and update guidelines throughout the research period. Because of the comprehensive instrumentation we have available, and our successful testing track record, we will be able to identify design considerations and present strong quantitative and qualitative evidence to support our final conclusions. At the time this paper was written, preliminary results were not yet available.

5.0 BIBLIOGRAPHY

- [1] Ausink, John A., William W. Taylor, James H. Bigelow, and Kevin Brancato. *Investment Strategies for Improving Fifth-Generation Fighter Training*. Santa Monica, CA: RAND Corporation, 2011.
http://www.rand.org/pubs/technical_reports/TR871. Also available in print form.
- [2] Lechner, R. and J. Wokurka. *Advancements of Integrated LVC Applied to Tactical Aviation Aircraft Training*. In Proceedings of the Interservice/Industry Training, Simulation and Education Conference (IIITSEC), 2010.
- [3] Möller, Björn, et al. *Towards Multi-Level Security for NATO Collective Mission Training – a White Paper*. In Proceedings of 2011 Spring Simulation Interoperability Workshop, 11S-SIW-069, SISO, April 2011.
- [4] Office of Naval Research. *Live, Virtual and Constructive (LVC) Training Fidelity*. Office of Naval Research Broad Agency Announcement (BAA) 11-005. Available online at www.fbo.gov/index?s=opportunity&mode=form&id=d28ec18fd28e5abf1df9e4858fed2373&tab=core&_cview=1
- [5] Schnell, Tom, James Melzer, and Steven J. Robbins. *The cognitive pilot helmet: enabling pilot-aware smart avionics*. In Head- and Helmet-Mounted Displays XIV: Design and Applications. 2009. Orlando, FL: SPIE.
- [6] Schnell, Tom, et al. *The Quality of Training Effectiveness Assessment (QTEA) Tool Applied to the Naval Aviation Training Context*. In Foundations of Augmented Cognition. Neuroergonomics and Operation Neuroscience. D. Schmorrow, I. Estabrooke, and M. Grootjen, Eds. 2009, Springer Berling/Heidelberg, pg. 640-649.
- [7] Schnell, Tom, Alex Postnikov, and Nancy Hamel. *Neuroergonomic Assessment of Simulator Fidelity in an Aviation Centric Live Virtual Constructive (LVC) Application*. In HCI International 2011. Springer: Hilton Orlando Bonnet Creek, Orlando, FL, USA.

3.11 Physiological Based Simulator Fidelity Design Guidance

PHYSIOLOGICAL BASED SIMULATOR FIDELITY DESIGN GUIDANCE

Tom “MACH” Schnell, Operator Performance Laboratory (OPL)

thomas-schnell@uiowa.edu

Nancy Hamel, Col. USAF (Ret), Advanced Infoneering, Inc.

nahamel@advancedinfoneering.com

Alex Postnikov, and Jaclyn Hoke, Rockwell Collins, Inc.

apostnik@rockwellcollins.com, jahoke@rockwellcollins.com

Angus “MAC” L. M. Thom McLean, III., ApotheoSys, Inc.

thom@apotheosys.com

Abstract. The evolution of the role of flight simulation has reinforced assumptions in aviation that the degree of realism in a simulation system directly correlates to the training benefit, i.e., more fidelity is always better. The construct of fidelity has several dimensions, including physical fidelity, functional fidelity, and cognitive fidelity. Interaction of different fidelity dimensions has an impact on trainee immersion, presence, and transfer of training. This paper discusses research results of a recent study that investigated if physiological-based methods could be used to determine the required level of simulator fidelity. Pilots performed a relatively complex flight task consisting of mission task elements of various levels of difficulty in a fixed base flight simulator and a real fighter jet trainer aircraft. Flight runs were performed using one forward visual channel of 40° field of view for the lowest level of fidelity, 120° field of view for the middle level of fidelity, and unrestricted field of view and full dynamic acceleration in the real airplane. Neuro-cognitive and physiological measures were collected under these conditions using the Cognitive Avionics Tool Set (CATS) and nonlinear closed form models for workload prediction were generated based on these data for the various mission task elements. One finding of the work described herein is that simple heart rate is a relatively good predictor of cognitive workload, even for short tasks with dynamic changes in cognitive loading. Additionally, we found that models that used a wide range of physiological and neuro-cognitive measures can further boost the accuracy of the workload prediction.

INTRODUCTION

The earliest development of simulators were directed towards familiarizing a student pilot with the basics of flight before the “real” instruction began in the aircraft. As the sophistication of aircraft and flight simulators increased, the utility of cockpit procedure trainers (CPTs) became clear – aircrew could learn the layout and function of the cockpit systems with greater efficiency on the ground than in the air. The incorporation of motion and visual systems allowed students to be introduced to even more sophisticated tasks and maneuvers prior to attempting them in the air. For certain tasks, such as delivering special ordinance or dealing with catastrophic systems failures, simulators provided the only way to prepare a pilot. This reliance on increasingly sophisticated simulators led to a belief that, in general, simulators should replicate as many aspects of the real-world as possible. For commercial aviation, this approach has produced simulators that can be certified to provide complete initial and recurrent aircrew qualification. Today, commercial aircraft flight simulators are

considered a cornerstone of flight instruction and pilot certification. The development of modern simulators has evolved to embrace the most sophisticated dimensions of flight training and mission rehearsal.

In flight simulators, physical fidelity relates to the accuracy of the physical layout of the crew station and how closely the visual, auditory, haptic, vestibular, and flight dynamic stimuli mimic those that will be experienced in the real aircraft. Functional fidelity primarily relates to how accurately the simulated crew station equipment acts like the operational equipment and cognitive fidelity is a quantification of how closely the human factors effects of the virtual environment track with those that will be found when operating in the real aircraft.

Statement of Problem

Designers of virtual environments for mission readiness training are ill-equipped to deal with difficult cost-benefit trade-offs that may affect fidelity and training effectiveness or transfer of training. Interaction of different fidelity dimensions

has an impact on trainee immersion, presence, and buy-in [1,2]. Flight simulators need to be designed to achieve a specific training objective. If too much simulator fidelity is specified for the objective, the simulator will be too expensive to acquire, operate, and maintain. If a device with too little fidelity is specified, the training objectives may not be met which may leave aircrew ill-prepared to perform the real-world flying tasks, giving rise to the possibility of loss of hull and life. Furthermore, simply specifying a very high fidelity in a flight simulator does not guarantee that all training objectives can be met with such a device.

What is needed is a quantitative method to characterize and predict the effects of fidelity on pilot performance. Using this method, flight simulator designers can then perform cost-benefit trade-offs and flight training organizations can adjust the training tasks to achieve the selected training objectives with the least amount of expenditure.

Solution Approach

In this paper, we describe the use of the Cognitive Avionics Tool Set (CATS) to quantify the effects of simulator fidelity on neuro-cognitive and physiological patterns exhibited by pilots. CATS was developed by the Operator Performance Laboratory (OPL) under a NASA aviation safety project entitled "Operator State Sensor Investigations" with additional funding provided by the Office Of Naval Research (ONR) and industry [3-5]. Our hypothesis is that a simulator of the highest fidelity will generate a human neuro-cognitive and physiological response that is indistinguishable from the one observed in the real aircraft.

In addition to quantifying the effects of simulator fidelity itself, we hypothesize that CATS could be used to design flight simulator curricula that challenge pilots to such an extent that their neuro-cognitive and physiological responses (arousal levels)

equal the intensity of the responses observed in the real airplane. We believe that training effectiveness in flight simulators depends on arousal levels that equal the ones observed in the real aircraft.

Apparatus

We developed a flight hardened human state sensor system for use on pilots flying real aircraft and flight simulators. We developed the necessary software to read the sensor state data, synchronize multiple channels into one stream of data, remove artifacts that arise at the sensor level, perform signal processing to characterize operator workload, and provide output to users in real-time and for after action review. The primary components of the overall software architecture is called the Cognitive Avionics Tool Set (CATS)[3, 4, 6].

In our experiment, the OPL L-29 research aircraft (Figure 1) and a matching flight simulator (Figure 2) provided the task stimulation at various levels of fidelity to the pilot who was fitted with a number of neuro-cognitive and physiological sensors including, electroencephalogram (EEG), electrocardiogram (ECG), respiration, and eye gaze tracking. As the pilot is performing a flight mission he/she responds to the cognitive demands of different mission task elements with characteristic neuro-cognitive and physiological patterns that are picked up by the sensors. The CATS processor continually reads the stream of data from the sensors, synchronizes those data into a unified system state vector, removes sensor artifacts, and performs signal processing such as fast Fourier transforms to provide the CATS operator state classifier with the data to calculate operator workload in real-time. This workload estimate is then transmitted to the plug-in of the instructor operator station which represents the primary user interface used by an instructor to monitor the training progress of a student. If desired, an instructor can open up CATS in a separate window to drill down into the data to investigate readings from individual sensors.

The experiment was performed on the OPL instrumented L-29 jet training aircraft (Figure 1) and a fixed base flight simulator (Figure 2). The content of the head-up and head-down displays in the flight simulator and the jet was identical. The aircraft was instrumented with an evaluation cockpit in the rear seat, integrated range instrumentation pods, a ground support infrastructure, and an operator monitoring and evaluation system. The aircraft was interconnected to a ground station using a range instrumentation datalink that can transmit in several formats, including the Advanced Range Data System (ARDS) protocol. This data link allowed for remote control of the experimental apparatus from our command and control trailer.



Figure 1. OPL's Instrumented L-29 Jet Trainer Aircraft



Figure 2. Fixed Base Flight Simulator

The flight simulator featured three channels of outside visuals, subtending a total of 120° lateral visual field of view (FOV) or around 40° per channel and a vertical field of view of 25 degrees. The outside visual (OSV) channels 2 and 3 were used to manipulate

the fidelity of the flight simulator with low fidelity corresponding to the condition where OSV 2 and 3 were off and medium fidelity when OSV 2 and 3 were on. The high fidelity level corresponded to runs in the L-29 jet.

Both the HDD and the HUD in the jet subtend a lateral field of view of 22°. The symbology and imagery that was generated on the HUD was conformal to this field of view. This means that the HUD was a conformal window into the real world. From the crew station in the rear cockpit, the evaluation pilot had unrestricted visibility of the real world except for the HUD, which provided a conformal inset of OSV symbology.

Figure 3 (left) shows the front cockpit of the L-29 jet where the safety pilot (SP) operated. The SP performed all maneuvering on the ground, take-off, landing, and repositioning of the aircraft between runs. The SP used standard aircraft instruments to navigate in US airspace under FAR part 91 flight rules. Two VHF radios were available to allow the SP to simultaneously communicate with air traffic control (ATC) and the command and control ground station on separate frequencies. A side display touch screen called the Phase Tagger was available to the SP to start and stop the recorder, tag events to check the video data link integrity, and to check CATS and the integrity of the eye tracker.

The rear cockpit (Figure 3, right) was the crew station that the evaluation pilot (EP, experiment participant) occupied. A daylight readable 15 inch touch screen display was installed in the head-down position allowing presentation of any avionics symbology as per program requirements. The symbologies could be driven either with PC board dedicated avionics graphics processors. In this experiment, the symbologies were identical to the ones used in the simulator and represented an instrument panel. A daylight readable 15

inch touch screen in the head-up display (HUD) position provided the same outside visuals as the center display in the simulator. The lateral FOV of the HUD display was 22° which made the imagery displayed on it conformal with the real world. Therefore, a pilot in the rear crew

station had an essentially unrestricted view of the surroundings, with the central 22° being a computer generated photorealistic inset and the remaining view being the real world.

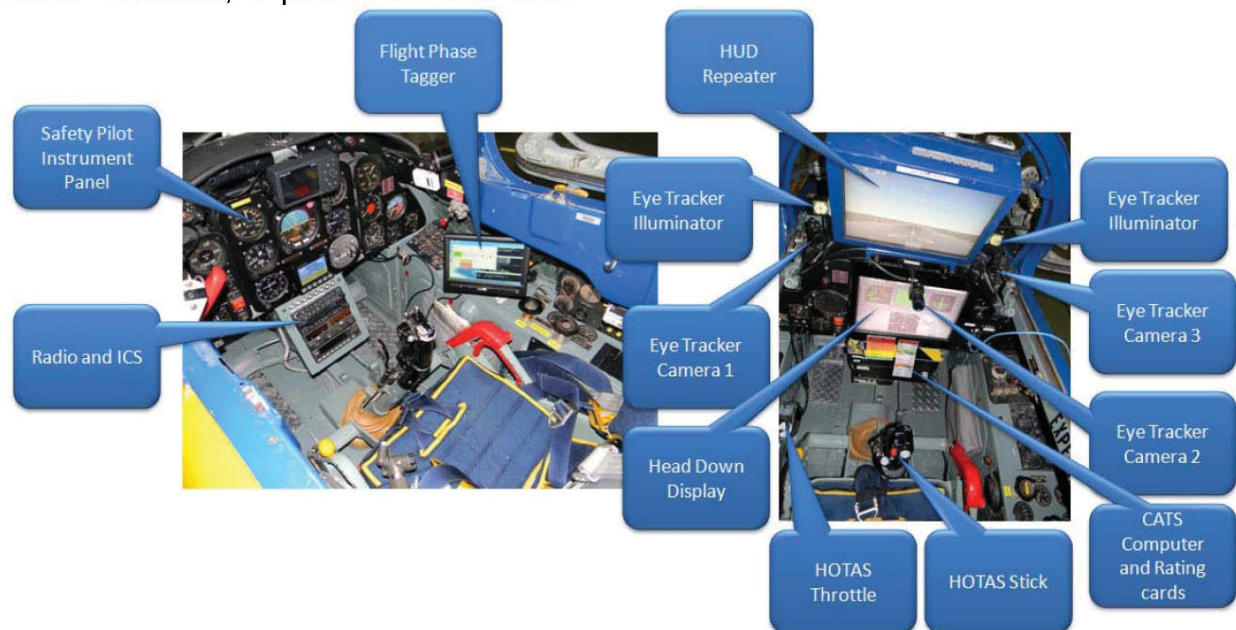


Figure 3. Front and Aft Cockpit of the L-29

To simplify the deployment of the neuro-cognitive and physiological sensors on the pilot we integrated the EEG electrodes in the liner of a flight helmet. The respiration belt and ECG electrodes were worn under the flight suit connecting to the peripheral electronics that are integrated in a pilot survival vest as shown in Figure 4. This level of integration provided for a ruggedized instrumentation package with a single point umbilical connection to the aircraft or flight simulator. The vest was designed from an ergonomics and flight safety point of view. The active electrode sensor cap was worn under the helmet with an additional soft helmet liner for wearer comfort. The EEG sensors were powered with batteries, which eliminated the need for an isolation transformer to ensure subject electrical safety.



Figure 4. Sensor Vest

One nice feature of the ActiCap electrode system is that each electrode has an LED, indicating the quality of the impedance of the sensor to skin interface. Before the experimenters put on the helmet over top of the sensor cap, a quick check of the LED status was performed to ensure acceptable impedance.

Figure 5 shows the overall data collection and analysis architecture that was used for this study. The figure illustrates the connections from the CATS sensor vest to the CART protocol processor. The CATS software can be run in real-time mode during data collection to provide the user with a detailed view of signal quality. However, running CATS during real-time is not necessary for experimental data collection, provided that the experimenters ensure that good quality data is being collected with CART. The CART GUI (sometimes referred to as the phase tagger) provides the experimenter with the ability to start and stop data collection for a particular run, or to advance the data collection indexed the next run. The recorded run index provides the user of CATS with the ability to query individual maneuvers during after action review. After the flight, the log files are extracted from the aircraft using either the datalink or a hardwired Ethernet hookup. The experimenter can then use the CATS output generator to create Google Earth files in the Keyhole Markup Language (KML) format of each run, with additional data overlays. This graphical depiction of the quality of the flight, including flight technical accuracy is an excellent way to perform after action review with students. For researchers, this tool is also an excellent way to determine root causes of strange performance effects. CATS generates detailed flight and mission performance plots and data sets that can be used for analysis of the accuracy of the flight. Batch mode processing in CATS generates summary data of the dependent variables for each run. This data can then be imported in additional packages such as Excel, Minitab, or Eureka for analysis.

Figure 5 illustrates how we used the Eureka data mining tool to generate workload equations for each of the 15 participants. These equations can be directly pasted into the expression parser of the workload gauges of CATS for subsequent real-time assessment of individualized workload.

EXPERIMENTAL DESIGN

We collected data in the OPL L 29 fighter jet trainer that served as the maximum fidelity upper baseline. Neuro-cognitive and physiological patterns observed under this maximum fidelity upper baseline were compared to the patterns observed in the flight simulator using a within subjects repeated measures experimental design.

We selected experimental tasks that included several flight maneuvers such as a simple climb, a holding pattern at a fix, a route with waypoint altitude clearance limits and speed assignments, and a wing-over maneuver that included several parameters such as pull-up point pulldown point, pitch angles, and bank angles, and an instrument approach (data not yet analyzed).

We used the number of variables that the pilot had to control for each maneuver as an indicator of the difficulty level of that maneuver as shown in Table 1.

Table 1. Assignment of Maneuver Difficulty to Task

Maneuver	Assumed Difficulty Level
Climb	1
Hold	2
Route	4
Wing Over	6

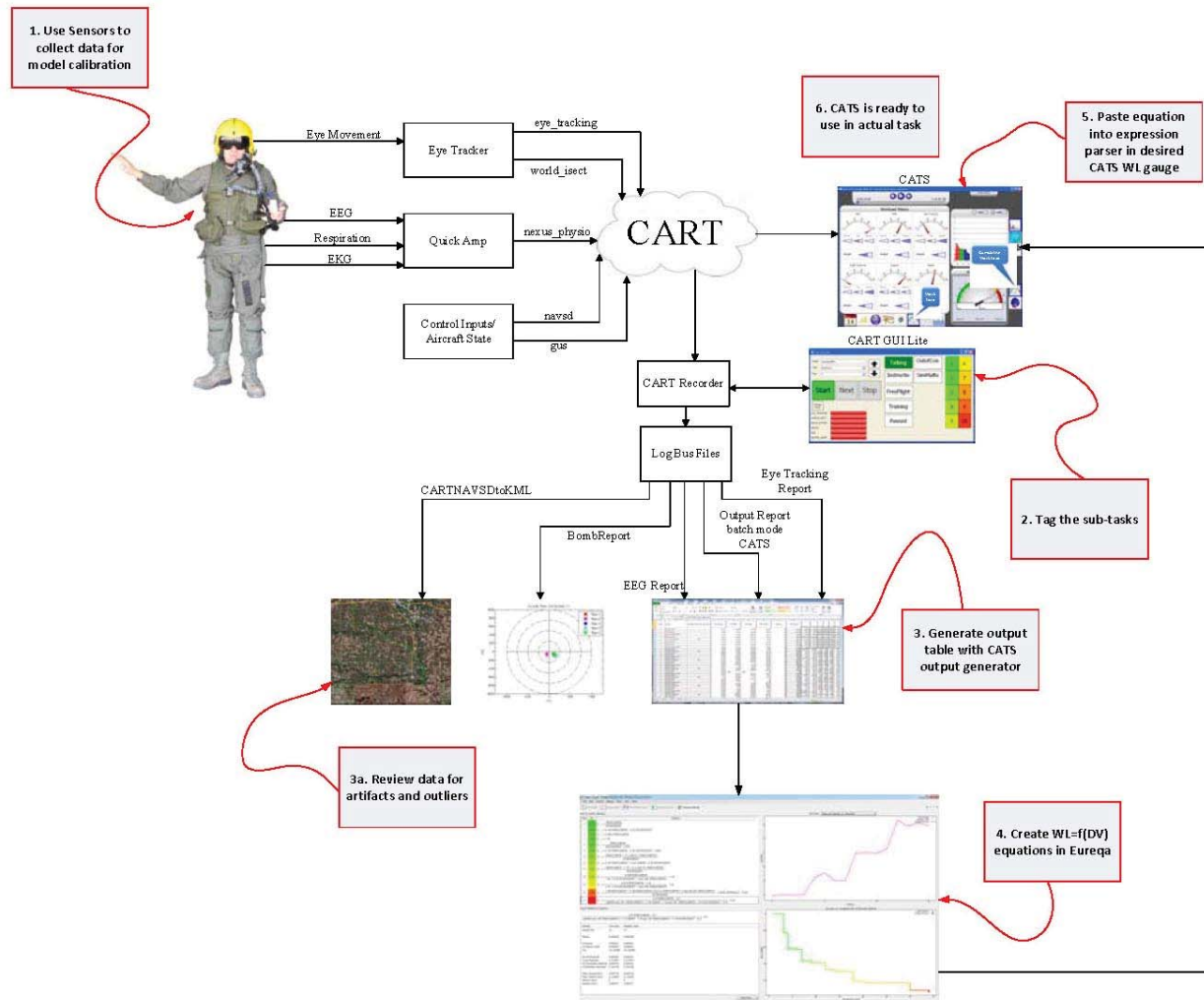


Figure 5. CART and CATS Data Collection and Analysis System

The dependent measures in CATS comprised in excess of 380 variables derived from EEG, ECG, eye movement, and flight technical performance. The EEG metrics consisted of power values in different clinical bands (delta, theta, alpha, beta, and gamma) for different scalp topographical regions (EOG, midline, sensorimotor, occipital), and different summary statistics (average, minimum, maximum, RMS, standard deviation). The ECG metrics consisted of straight heart rate, the percentage of heartbeats outside of a 20 ms or 50 ms window, and short-term heart rate trend. Eye fixation metrics included the number of fixations in areas of interest, fixation duration, and distance between subsequent fixations. There are

many task specific flight technical measures that were used including cross track error, vertical track error, speed error, climb angle and descent angle.

A total of 15 subjects were used in this study. This pool of subjects was divided into three groups of five each. Group 1 consisted of those pilots who flew the simulator using only one channel of outside visuals, Group 2 flew the simulator with three outside visual channels, and Group 3 flew their mission in the real aircraft. Each pilot participant performed five replications of each run.

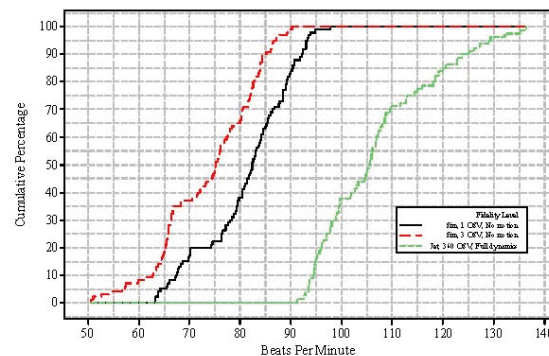
RESULTS

Effect of Fidelity on Selected Variables

In this paper we will focus on the individualized operator state models. Only a small selection of physiological data is discussed herein. Additional results can be found in [2]. Figure 6 shows the effect of simulator fidelity on heart rate using a cumulative histogram. The leftmost trace indicates a heart rate distribution for those pilots who flew their mission in the simulator using three channels of outside visuals with no motion. The three outside visual channels provided the pilot with sufficient lateral visual information to allow performance of the maneuvering tasks with relative ease. This is reflected in a lower heart rate. With no change other than turning off the two peripheral outside visual displays and flying the mission with only one channel of outside visuals, the heart rate increased statistically significantly. This is a clear indication that the pilots had to compensate for the lack of field of view during the maneuvering tasks using additional cognitive resources in the process. The increase in simulator fidelity reduced the level of cognitive loading because more of the necessary information is available in the three channel simulator condition.

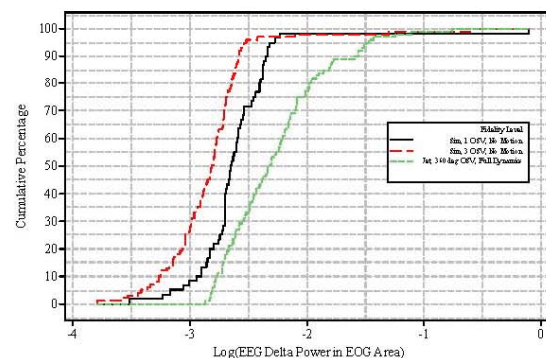
One could argue that the flights performed in the real aircraft should result in even less cognitive workload and arousal as the visual display is volumetric and 360° at real-world resolution. However, this is not what we found. Instead, we found that the heart rate was significantly higher in the real aircraft when compared to both simulator conditions. This, of course, makes complete sense as not only the visuals change, but also the motion dynamics, acoustics, and the knowledge that one is in fact in a real airplane flying at high speeds and high altitudes. Pilots flying in the fixed base simulator did not feel the effects of acceleration, whereas pilots in the aircraft did. We saw many more turn overshoots in

the aircraft when compared to the simulator. The pilots who were participated in this study had not flown nimble acrobatic jet aircraft before. These participants were therefore highly aroused in anticipation of the mission on hand. Flying the aircraft also provided additional stimulation in the olfactory, auditory, and proprioceptive stimulus dimensions. Therefore, we feel that the heart rate increase observed for the aircraft sorties in Figure 6 are consistent with our expectations. A similar effect was found in EEG (Figure 7).



N=15 participants at 5 replications. ANOVA: $F_{3,42}=269.06$, $p<0.001$. Pairwise T: Level 1 vs. Level 2 $T=-5.395$, $p=0.00001$, Level 1 vs. Level 3 $T=17.354$, $p=0.00001$, Level 2 vs. Level 3 $T=22.44$, $p<0.00001$

Figure 6. Beats per Minute Effect for Simulator Fidelity Levels

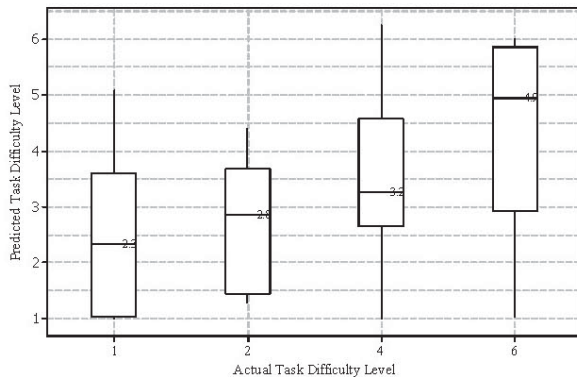


N=15 participants at 5 replications. ANOVA: $F_{3,42}=49.35$, $p<0.0001$. Pairwise T: Level 1 vs. Level 2 $T=-3.442$, $p=0.0017$, Level 1 vs. Level 3 $T=5.123$, $p=0.00001$, Level 2 vs. Level 3 $T=9.891$, $p<0.00001$

Figure 7. Log of EEG Power in EOG Area for Fidelity

Personalized Workload Models

We used the study data to build individualized workload equations, separately for each participant. We used replications 1,2,4, and 5 to build the workload model and we validated the resulting equations against data from replication 3. This sequence was chosen because we believe that replications 1 and 2 were in the steep part of the learning curve and replications 4 and 5 were in the flat part of the learning curve. Combined, this training data set represented a fairly neutral learning condition. The Replication 3 validation data point was thought to be in the neutral region between the steep and flat part of the learning curve. A total of 15 equations (Table 3) were generated using Eureqa and entered into Excel for validation against the third application data set. The resulting predicted task difficulty level was then plotted against the actual task difficulty level as shown in Figure 8. The correlation between the actual and predicted data set was 0.888. We feel this is a fairly respectable result in terms of accuracy of prediction.



Correlation between actual and predicted data is 0.888

Figure 8. Predicted vs. Actual Task Difficulty Level using Neurocognitive and Physiological Measures

The workload models shown in Table 3 were developed by using Eureqa to determine the significant variables to construct a nonlinear regression model for

each participant. An explanation of the variable names is given in Table 2

Table 2. Main Variables used in Workload Prediction Models

Variable	Meaning
PERCLOBPM	Percent difference of task heartbeat compared to rest
LPNN20	Log percentage of heartbeats falling outside of a 20 ms inter-beat interval
LBLM	Log beta low midline EEG power
LBO	Log beta occipital EEG power
LDS	Log delta sensory motor EEG power
LDM	Log delta midline EEG power
HUDFIXDIST	Distance between subsequent eye fixations to the HUD
HUDFDMEAN	Mean duration of eye fixations made to the HUD

DISCUSSION

Four flight tasks of increasing levels of difficulty were used in 15 pilots to elicit a range of workload responses. Neuro-cognitive and physiological measures were collected for three levels of simulator fidelity (low, medium, high) using the Cognitive Avionics Tool Set (CATS). Nonlinear closed form models for workload prediction were generated based on these data for the various mission task elements

Our data indicate that there is a significant difference in the neuro-cognitive and physiological patterns obtained from pilots flying a fixed base flight simulator when compared to flying the real aircraft. We also found that the real aircraft provided considerably more cognitive arousal to the pilots used in our experiment when

compared to the level of arousal that was provided by the flight simulator. We discovered that factors such as the extent of the field of view of the outside visuals in a flight simulator can considerably affect neuro-cognitive and physiological patterns which can be discriminated with the algorithms integrated in CATS.

Based on our results, it seems that simple heart rate is a relatively good predictor of cognitive workload, even for short tasks with dynamic changes in cognitive loading. Additionally, we found that models which use a wide range of physiological and neuro-cognitive measures can further boost the accuracy of the workload prediction. By comparing the physiological and neuro-cognitive response of pilots in a simulated environment with the corresponding response in real aircraft, it was possible to generate measures of performance that track with the fidelity of a virtual environment.

Based on our data, we can conclude with confidence that pilots flying the aircraft struggled more and generated larger flight technical errors when compared to pilots flying the simulator. However, the percentage of improvement from the first to the fifth replication was much larger in the aircraft than it was in the flight simulator. We believe that the multisensory stimulation in the aircraft aids pilots in improving flight control inputs from one replication to the next. Operators of fixed base training flight simulators should consider increasing task difficulty beyond what is expected in the real aircraft. This increase in task difficulty will increase the trainee's arousal level thereby generating neuro-cognitive and physiological engagement levels approaching those observed in the real aircraft.

This study is limited to one incarnation of a fixed base flight simulator and a study with a relatively small sample size. Without performing a study in a high-end motion base simulator, we simply cannot

hypothesize on the performance that one would observe in a high-end simulation device. However, it appears that the multisensory stimulation experienced in the real aircraft aids the pilot in refining their sensorimotor skills throughout about five replications.

REFERENCES

1. Alexander, A., L. Brunye, and T. Weil. *From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games*. 2005; Available from: <http://www.darwars.org/downloads/DARWARS%20Paper%2012205.pdf>.
2. Schnell, T., A. Postnikov, and N. Hamel, *Neuroergonomic Assessment of Simulator Fidelity in an Aviation Centric Live Virtual Constructive (LVC) Application*, in *HCI International 2011*. 2011, Springer: Hilton Orlando Bonnet Creek, Orlando, Florida, USA.
3. Schnell, T., M. Keller, and T. Macuda, *Application of the Cognitive Avionics Tool Set (CATS) in Airborne Operator State Classification*. Augmented Cognition International Conference, 2007.
4. Schnell, T., M. Keller, and P. Poolman. *Neurophysiological workload assessment in flight*. in *Digital Avionics Systems Conference, 2008. DASC 2008. IEEE/AIAA 27th*. 2008.
5. Schnell, T., T. Macuda, and M. Keller, *Sensor Integration to Characterize Operator State*, in *Augmented Cognition; A Practitioner's Guide*, D.D.S.K.M. Stanney, Editor. 2008, Human Factors and Ergonomics Society. p. 33.
6. Schnell, T., J.E. Melzer, and S.J. Robbins. *The cognitive pilot helmet: enabling pilot-aware smart avionics*. in *Head- and Helmet-Mounted Displays XIV: Design and Applications, April 16, 2009 - April 16, 2009*. 2009. Orlando, FL, United states: SPIE.

Table 3. Task Difficulty Prediction Models for CAS Task

Subject	Task Difficulty=f(Dependent Variables)	R ²	Fitness	
1	$1.3184637 \cdot \cos(\text{PERCLOBPM} + 1.3555878 \cdot \text{LBLM}) + 0.12370914 \cdot \text{PERCLOBPM} + 0.51329815 \cdot \text{LBLM} + 1.6968185 \cdot \sin(221.40964 \cdot \text{LBLM}) - 1.1071174 \cdot \sin(2.0060091 \cdot \text{PERCLOBPM}) - 0.059559464$	0.991	0.059	Simulator, 1 C channel of OSV
2	$1.0827115 + (3.3491788 \cdot \text{LPNN20} + 0.28201574 \cdot \text{PERCLOBPM} + \sin(69.19706 \cdot \text{LBO}) + 2.353423$	0.96	0.149	
3	$0.84847742 + (0.16712549 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM} + -0.91743886 \cdot \text{HUFDXIDIST} \cdot \cos(-9.3917542 \cdot \text{PERCLOBPM}) \cdot \cos(-9.3917542 \cdot \text{PERCLOBPM}) / \text{HUFDMEAN} - 25.426651 \cdot \cos(-9.3917542 \cdot \text{PERCLOBPM}) \cdot \cos(-9.3917542 \cdot \text{PERCLOBPM}) \cdot \cos(\text{PERCLOBPM} - 1.4302194 \cdot \text{HUFDMEAN})) / \text{HUFDXIDIST}$	0.99	0.061	
4	$1.564743 + \text{HUFDMEAN} + 0.34546053 \cdot \text{PERCLOBPM} + 1.2832386 \cdot \cos(\text{LDM} \cdot \text{LBLM}) + 0.37645277 \cdot \cos(1.1469964 \cdot \text{PERCLOBPM}) \cdot \cos(1.4169006 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM}) + 2.9303737 \cdot \text{HUFDMEAN} \cdot \cos(1.1469964 \cdot \text{PERCLOBPM}) - 0.46708772 \cdot \cos(1.4169006 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM})$	0.996	0.045	
5	$2.7095852 + \cos(\text{PERCLOBPM}) \cdot \cos(\text{PERCLOBPM} - 4.4980625e-5) / (0.11290019 + 1.5396878 / (\text{PERCLOBPM} - 4.4980625e-5)) + 0.91720986 \cdot \cos(\text{PERCLOBPM}) - 0.75311863 \cdot \cos(\text{PERCLOBPM} - 4.4980625e-5) \cdot \cos(2.763207 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM}) - 0.045648079 / \sin(\text{PERCLOBPM}) - 3.929251 \cdot \text{HUFDMEAN}$	0.991	0.049	
6	$5.8512836 + 2.6331587 \cdot \cos(10.289939 \cdot \text{PERCLOBPM}) / (\text{PERCLOBPM} - 3.6436858) + (2.5140517 \cdot \text{HUFDXIDIST} \cdot \sin(\sin(14.251148 \cdot \text{LBO})) - 22.866884 \cdot \sin(\sin(14.251148 \cdot \text{LBO}))) / (\cos(\text{PERCLOBPM} - 3.6436858 / \text{PERCLOBPM}) + \text{PERCLOBPM} - 4.1335578)$	0.995	0.041	Simulator, 3 C channels of OSV
7	$2.609494 + \sin(0.20168348 \cdot \text{PERCLOBPM}) + 0.20168348 \cdot \text{PERCLOBPM} \cdot \text{HUFDMEAN} \cdot \sin(0.20168348 \cdot \text{PERCLOBPM}) \cdot \sin(-3.2543337 \cdot \text{LBLM} / \text{LPNN20}) \cdot \sin(-3.2543337 \cdot \text{LBLM} / \text{LPNN20}) + 11.91309 \cdot \text{HUFDMEAN} \cdot \text{HUFDMEAN} \cdot \text{HUFDMEAN} \cdot \sin(-3.2543337 \cdot \text{LBLM} / \text{LPNN20}) +$	0.975	0.081	
8	$0.27726284 + 5.3854728 \cdot \cos(\text{LTS}) + 5.3854728 \cdot \text{HUFDMEAN} \cdot \cos(\text{HUFDMEAN}) + 0.64272124 \cdot \sin(-2959.6062 \cdot \text{LBO}) + 1.5181508 \cdot \sin(-0.17662553 \cdot \text{HUFDXIDIST}) / (\sin(-626.79443 \cdot \text{LBO}) - 0.965011)$	0.997	0.036	
9	$2.8691647 + \sin(4.32651 \cdot \text{LDS} + 2.6206877 \cdot \text{PERCLOBPM} \cdot \text{LDS}) + 2.6206877 \cdot \text{LDS} \cdot \cos(\text{LDS} \cdot \text{LDM}) + 9.831584 \cdot \cos(\text{LDS} \cdot \text{LDM}) - 3.0779896 \cdot \log(\text{PERCLOBPM}) \cdot \cos(\text{LDS} \cdot \text{LDM}) - 0.58662623 / \text{LDS} - 0.14015104 \cdot \text{PERCLOBPM}$	0.97	0.104	
10	$3.2809136 + 1578.9419 / (-1252.537 \cdot \text{LDM}) + 2.6978884 \cdot \cos(106.14892 \cdot \text{LDM}) + 0.68518925 \cdot \cos(-1252.537 \cdot \text{LDM}) + 3.3517282 \cdot \sin(108.99526 \cdot \text{LDM}) - 1.4094001 \cdot \cos(196.23761 \cdot \text{LDM}) \cdot \cos(106.14892 \cdot \text{LDM})$	0.993	0.044	
11	$3.2083549 + 0.31052729 \cdot \cos(231.55759 \cdot \text{HUFDXIDIST}) + 2.2928617 \cdot \cos(-10.254387 \cdot \text{HUFDXIDIST}) + 0.22617097 \cdot \text{HUFDXIDIST} \cdot \cos(-18.568399 \cdot \text{LDM}) \cdot \cos(-10.222954 \cdot \text{HUFDXIDIST} \cdot \text{HUFDXIDIST}) + (3.2083549 + 0.31052729 \cdot \cos(231.55759 \cdot \text{HUFDXIDIST}) + 2.2928617 \cdot \cos(-10.254387 \cdot \text{HUFDXIDIST}) +$	0.998	0.024	Aircraft
12	$4.3274031 + 2.0209646 \cdot \sin(0.13923776 \cdot \text{PERCLOBPM}) + 2.3252544 \cdot \sin(0.12731959 \cdot \text{PERCLOBPM}) \cdot \sin(2.7443035 \cdot \text{PERCLOBPM}) - 1.5827621 \cdot \sin(0.12731959 \cdot \text{PERCLOBPM}) \cdot \sin(2.2408776 \cdot \text{PERCLOBPM}) \cdot \sin(2.7443035 \cdot \text{PERCLOBPM}) - \sin(2.2408776 \cdot \text{PERCLOBPM}) \cdot \cos(\sin(0.13923776 \cdot \text{PERCLOBPM}))$	0.98	0.058	
13	$1.2922051 \cdot \sin(103.67542 / \text{LPNN20}) + 0.13030657 \cdot \text{PERCLOBPM} + 0.026240172 \cdot \text{PERCLOBPM} / \text{LPNN20} + \cos(\sin(103.67542 / \text{LPNN20})) - 1.7087533 \cdot \sin(56.959488 / \text{LPNN20}) \cdot \cos(0.13030657 \cdot \text{PERCLOBPM} + 1.2922051 / (0.13030657 \cdot \text{PERCLOBPM} + 0.026240172 \cdot \text{PERCLOBPM} / \text{LPNN20})) - 0.72068655$	0.994	0.058	
14	$(-2.6239846 \cdot \text{PERCLOBPM} - 217.39574) / (\text{LDEOG} \cdot \cos(-386.85562 \cdot \text{PERCLOBPM}) + 3.2971733 \cdot \text{LDEOG} - 1.7448196 \cdot \cos(-587.14429 \cdot \text{PERCLOBPM}) - 0.57358128 \cdot \text{HUFDXIDIST} - 11.827497) - 8.500103$	0.99	0.068	
15	$0.0022779282 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM} + 3.7092943 \cdot \text{LDS} + 2.5593133 \cdot \text{HUFDMEAN} - 7.2671909 \cdot \text{LTS} - 1.4026718 \cdot \log(0.0019029471 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM}) \cdot \sin(4.7368174 \cdot \text{PERCLOBPM} \cdot \text{PERCLOBPM}) - 23.494602$	0.988	0.072	

3.12 Predicting the Impacts of Intravehicular Displays on Driving Performance with Human Performance Modeling

Predicting the Impacts of Intravehicular Displays on Driving Performance with Human Performance Modeling

Diane Kuhl Mitchell, Josephine Wojciechowski, and Charneta Samms
Army Research Laboratory, Human Research & Engineering Directorate
Aberdeen Proving Ground, Maryland
diane.k.mitchell.civ@mail.mil; josephine.g.wojciechowski.civ@mail.mil;
charneta.l.samms.civ@mail.mil

Abstract. A challenge facing the U.S. National Highway Traffic Safety Administration (NHTSA), as well as international safety experts, is the need to educate car drivers about the dangers associated with performing distraction tasks while driving. Researchers working for the U.S. Army Research Laboratory have developed a technique for predicting the increase in mental workload that results when distraction tasks are combined with driving. They implement this technique using human performance modeling. They have predicted workload associated with driving combined with cell phone use. In addition, they have predicted the workload associated with driving military vehicles combined with threat detection. Their technique can be used by safety personnel internationally to demonstrate the dangers of combining distracter tasks with driving and to mitigate the safety risks.

1.0 INTRODUCTION

Driving is a behavior humans perform on a daily basis. Unfortunately, so is distracted driving. According to the definition developed by the US National Highway Traffic Safety Administration (NHTSA) distracted driving occurs when drivers divert their attention from the driving task to focus on other activities (2010). Using a cell phone to text or talk or navigating with an intravehicular GPS system are examples of common distracter tasks drivers combine with driving. Regrettably these distracting activities can have extremely adverse effects on driving performance.

In 2009, within the US, distracted driving contributed to 448,000 motor vehicle injuries and 5,474 deaths (National Highway Traffic Safety Administration, 2010). However, the issue of distracted driving is not limited to the U.S. In May 2010, recognizing the risks associated with distracted driving, U.N. Secretary General Ban Ki-moon released a directive that banned the 40,000 United Nations staff from texting while they are driving U.N.-owned vehicles (U.S. Department of Transportation, 2010). In addition, driving and distraction consequences are not limited to the civilian population. Motor vehicle accidents are the primary cause of fatalities among U.S. military personnel (Krahl, Jankosky, Thomas, & Hooper, 2010). 84% of the accidents involved lack of awareness with approximately half

attributed to distraction (Ecola, Collins & Eiseman, 2010).

While safety experts compile statistics demonstrating the risks associated with distracted driving, many types of within-vehicle devices are becoming inexpensive and readily available to drivers. It is critical, therefore, for safety personnel to educate drivers about best practices for using intravehicular devices and for system engineers to develop within-vehicle devices with designs that minimize driver distraction. Human performance modeling is an inexpensive and effective method for achieving these goals.

Human performance models can simulate drivers driving with and without distracter tasks (Horrey, Wickens, & Consalus, 2006). System designers can compare the driving performance from these simulations and select the intravehicular device with the design that has the least effect on driving performance. In addition, they can use the models to determine if the devices can be used while driving and, if so, the optimum technique for device operation during driving. Risk management tools can be employed using the risk severity and frequency of task combinations to evaluate safety. They can then educate drivers by incorporating the best technique within drivers' education and safety programs.

Recognizing the capabilities and benefits of a human performance model of

distracted driving, analysts at the Human Research & Engineering Directorate (HRED) of the Army Research Laboratory (ARL) used the Improved Performance Research Integration Tool (IMPRINT) to model driving.

IMPRINT is a dynamic, stochastic, discrete event simulation tool designed to predict system and mission performance to support system design (U.S. Army Research Laboratory, 2010). IMPRINT analysts build task network models representing the human interacting with a system to perform a specific mission. The mission is represented by various functions and tasks that are required to complete the mission. The analyst enters parameters for each task such as the time it takes to perform the task and how much mental demand the task will require of the operator. The functions and tasks are then linked and coded to create a stochastic task network. During the model execution, IMPRINT calculates the mental workload associated with the combination of all the tasks the operator is performing simultaneously over the entire mission and provides output reports for the analyst to review. Using these output reports, the analyst can examine the mental workload profile of each operator over the entire mission to identify points where workload is high which is an indicator of potential performance decrements.

2.0 DRIVER WORKLOAD MODEL

Using IMPRINT, (Wojciechowski, 2004) developed the initial driving model for ARL. She developed this Driver Workload Model (DWM) to have a representation of driving in IMPRINT that could be used to investigate the mental workload of driving concurrent with communications. She developed the goals, functions, and tasks for the physical characteristics of driving in the DWM using hierarchical task analysis methods (Kirwan & Ainsworth, 1992). She augmented these with cognitive task analysis (Cooke, 1994) to include the non-physical aspects of

driving and controlling vehicles. In the DWM, the driver's primary goal includes the main functions of driving; visual function (see), psychomotor function (move), and the cognitive function (maintaining situation awareness (SA). Communications is an additional function the driver performs (Fig. 1).

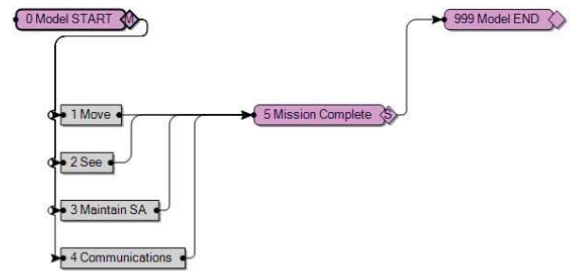


Figure 1 IMPRINT Driver Workload Model

The visual function of the DWM contained the tasks associated with visual processing of driving (Fig. 2). These tasks are; scanning the sector, detecting landmarks, recognizing the path, calculating the distance to the objective, and comparing to guidance or knowledge. Scanning the sector is the first task conducted. After scanning, the model uses a probabilistic decision to determine if a landmark is in the path. When a landmark is detected, two subsequent tasks are triggered simultaneously. They are recognizing the path, calculating the distance to the objective, and comparing to guidance or knowledge. The tasks in the visual function occur in a feedback loop similar to Wickens' human information processing model (Wickens and Hollands, 2000).

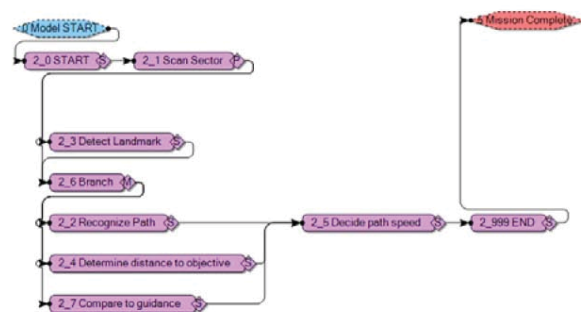


Figure 2 IMPRINT Driver Workload Model – Visual (See) Function

The psychomotor function contains tasks associated with the physical tasks of driving (Fig. 3). They are steering (and non-steering) and accelerating, decelerating, or coasting. As the model begins, there is an initial acceleration. Then a probabilistic decision is made to whether the driver accelerates, decelerates, or coasts. The model cycles through this loop to continually adjust the speed. Model coding can be used to influence this loop depending on the use of the DWM. Running simultaneously with the speed control loop is the directional control loop. This is a loop of two tasks; steer (adjust the steering mechanism) and do not steer (hold the steering mechanism steady).

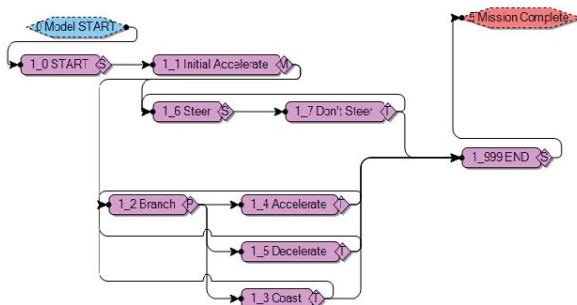


Figure 3 IMPRINT Driver Workload Model – Psychomotor (Move) Function

The cognitive function of driving was developed as a direct result of cognitive task analysis (Fig. 4). This function consists of tasks that represent assessing the orientation of the vehicle, assessing the motion of the vehicle, assessing the traction of the vehicle, and an awareness of the vehicle function. In typical driving scenarios, these tasks are quickly learned and automatic, based on cues from the environment. They do not require a lot of attentional demand or memory capacity as represented in Wickens' model (Wickens and Hollands, 2000). When the tasks are completed the model probabilistically decides if there is an issue with the vehicle. Then, the model cycles back through the cognitive tasks.

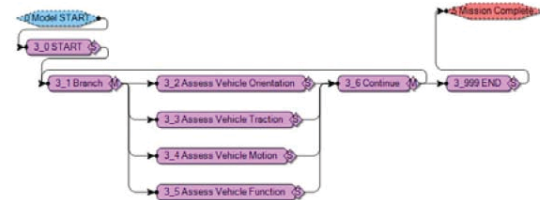


Figure 3 IMPRINT Driver Workload Model – Maintain Situation Awareness (SA) Function

The communication function in the DWM includes tasks representing communication with someone outside the vehicle (Fig. 5). They can be triggered randomly or designed to happen at specific intervals.

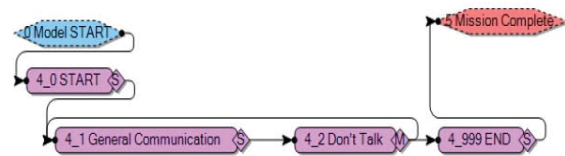


Figure 4 IMPRINT Driver Workload Model – Communication Function

In 2006, Wojciechowski conducted a study that verified the DWM accurately simulated driving performance and validated the workload predictions from the DWM. The study investigated the impact on performance and workload of driving while attending to an auditory task using both the DWM and a driving simulator. The results of the model and the simulator experiments were then compared to see if there was a correlation between the DWM workload predictions and the self-report workload ratings and performance of participants in the simulation experiment.

The results of both the DWM and the experiment indicate an impact of an auditory task on driving. The mental workload predictions from the DWM correlated with the self-report workload ratings from the experimental simulation. Therefore the DWM is a good predictor of mental workload changes with auditory secondary tasks. While the workload numbers correlate with simulator numbers, this simulator did not have the fidelity to

demonstrate what the workload changes would mean in terms of performance measures. Additional testing is necessary to demonstrate performance changes due to workload increases predicted by the model.

3.0 DISCUSSION

The DWM has the benefit of providing analysts with a representation of driving that is generic enough that it can be simplified or expanded to simulate driving within a number of contexts. It was created to investigate off-road driving under differing levels of autonomy (Wojciechowski, Kogler, & Lockett, 2001). The off-road driving model has been expanded by an ARL analyst to represent autonomous control of an unmanned asset within a military mission (Mitchell, 2005). In addition, when building multi-task models of military missions, ARL analysts have simplified the DWM to represent driving within workload analyses of military missions (Mitchell & Samms, 2009). Consistent with the workload predictions from the initial DWM, the workload predictions from these multi-tasks models have been verified in laboratory and field studies (Chen & Joyner, 2009); (McDowell, Nunez, Hutchins, & Metcalfe, 2008).

The workload predictions from the ARL analyses of driving consistently predicted high workload associated with driving. Concurrently, safety organizations such as the NHTSA, were releasing statistics demonstrating the adverse effects of distracter tasks on driving performance. Consistent with the ARL analyses and safety statistics, researchers in industry and academia demonstrated that these detriments to driving performance occur because driving is associated with high mental workload (Reimer, 2010); (Ranney T. , 2008); (Battelle, 2006); (Horrey, Wickens, & Consalus, 2006); (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006); (McCallum, Campbell, Richard, & Brown, 2006); (Horberry, Anderson, Regan,

Triggs, & Brown, 2005); (Horrey, Alexander, & Wickens, 2003); (Kurahashi, Motonori, & Akamatsu, 2003); (Ranney, Mazzae, Garrott, & Goodman, 2000); (Tijerina, 1995); (Green, Lin, & Bagian, 1993); (Wierwille W. , 1993); (Wierwille, Tijerina, Kiger, Rockwell, Lauber, & Bittner Jr., 1992).

Although the ARL analysts' modifications to the DWM model have been for military applications, it can be modified for use by academic or industry researchers as well. The DWM, for example, could be modified to focus exclusively on the visual function of driving.

Driving research studies are in agreement that driving is a predominantly visual activity. Therefore, when time-sharing monitoring intravehicular displays with driving, drivers will focus their eyes on the forward terrain and use a short glance to look at an intravehicular display. If they cannot get all the information they need with one glance, they will alternate multiple glances with glances at the road ahead.

To represent this driver visual behavior, researchers could modify the visual function in the DWM to represent the driver alternating glances at the forward terrain with glances at intravehicular displays such as speedometers, warning and advisory alerts or navigation systems. They could then use this modified version of the DWM to compare the number of glances required for alternative intravehicular display designs. The display with the minimal number of glances would be recommended as the optimum design. ARL analysts have successfully implemented this approach in an analysis of the driver of a military mine detection vehicle (Mitchell & Sweeney, in review).

Researchers could also use the DWM to develop guidelines for use of intravehicular displays while driving. For this analysis, they could develop several DWM models with drivers glancing away for varying

lengths of time. By comparing the driving performance from the alternative models they could establish thresholds for glance duration times.

Additionally, the DWM could be used to address safety issues with distracted driving using safety assessment methods (Swallom, Lindberg, & Smith-Jackson, 2003). Risk analysis measures can be used because IMPRINT provides frequency of task combinations and workload levels. Severity of the risk can be estimated by the workload level. Critical task flows can be used in fault tree analysis to determine potential causes of hazardous combinations of tasks. Additionally, event tree analysis of multiple model runs can predict probability of hazardous combinations of tasks.

4.0 CONCLUSION

The DWM demonstrates the feasibility of using human performance modeling to represent driving. The inclusion of the DWM into many military analyses illustrates its adaptability for representing a variety of driving applications. The IMPRINT software that ARL analysts used to develop the DWM is available to universities, government organizations, and government contractors. Therefore, the DWM provides an inexpensive technique for predicting the impacts of intravehicular displays on driver performance. The statistics supporting the adverse consequences of distracted driver verify the need for this cost-effective modeling technique.

5.0 REFERENCES

- Battelle. (2006). *Integrated Vehicle-Based Safety System Heavy Truck Driver Vehicle Interface (DVI) Literature Review*. Seattle, WA: Center for Human Performance and Safety.
- Chen, J. Y., & Joyner, C. T. (2009). Concurrent Performance of Gunner's and Robotics Operator's Tasks in a Multitasking Environment. *Military Psychology*, 21 (1), 98-113.
- Cooke, N. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41 (6), 801-849.
- Ecola, L., Collins, R. L., & Eiseman, E. (2010). Understanding and reducing off duty vehicle crashes among military personnel (ISBN 978-0-8330-5021-2). RAND National Defense Research Institute. available online. http://www.rand.org/pubs/technical_reports/2010/RAND_TR820.pdf. Accessed August 2, 2011.
- Green, P., Lin, B., & Bagian, T. (1993). *Driver workload as a function of road geometry: a pilot experiment*. Ann Arbor: University of Michigan Transportation Research Institute.
- Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2005). *Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance*. Australia: National Roads and Motorists' Association.
- Horrey, W. J., Alexander, A. L., & Wickens, C. D. (2003). Does Workload Modulate the Effects of In-Vehicle Display Location on Concurrent Driving and Side Task Performance? *Driving Simulation Conference North America 2003 Proceedings*. Dearborn, MI.
- Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling Drivers' Visual Attention Allocation While Interacting With In-Vehicle Technologies. *Journal of Experimental Psychology*, 12 (2), 67-78.
- Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London: Taylor and Francis.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data (DOT HS 810 594)*. Washington, DC: National Highway Traffic Safety Administration.
- Krahl, P. L., Jankosky, C. J., Thomas, R. J., & Hooper, T. I. (2010). Systematic Review of Military Motor Vehicle Crash-Related Injuries. *American Journal of Preventative Medicine*. 38(1S):S189–S196. Accessed online August 2, 2011 [http://phc.amedd.army.mil/PHC%20Resource%20Library/Systematic%20Review%20of%](http://phc.amedd.army.mil/PHC%20Resource%20Library/Systematic%20Review%20of%20)

- 20Military%20Motor%20Vehicle%20Crash-Related%20Injuries.pdf
- Kurahashi, T., Motonori, I., & Akamatsu, M. (2003). Objective measures to assess workload for car driving. *SICE 2003 Annual Conference*, 1, pp. 270-275. Fukui, Japan.
- McCallum, M., Campbell, J. L., Richard, C., & Brown, J. L. (2006). *Integrated vehicle-based safety system heavy truck driver vehicle interface (DVI) literature review (UMTRI-2008-22)*. Seattle, WA: Battelle Center for Human Performance and Safety.
- McDowell, K., Nunez, P., Hutchins, S., & Metcalfe, J. S. (2008). Secure mobility and the Autonomous Driver. *IEEE Transactions on Robotics*, 24 (3), 688-697.
- Mitchell, D. (2005). *Soldier Workload Analysis of the MCS Platoon's Use of Unmanned Assets (ARL-TR-3476)*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.
- Mitchell, D., & Samms, C. L. (2009). Workload Warriors: Lessons Learned from a Decade of Mental Workload Prediction Using Human Performance Modeling. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 819-823). San Antonio, TX: Human Factors and Ergonomics Society.
- Mitchell, D., & Sweeney, H. (in review). *Workload analysis of the operator of the Husky-mounted Mine Detection System*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.
- National Highway Traffic Safety Administration. (2010). Retrieved May 2011, from Distracted Driving: <http://www.distracted-driving.gov/research/PDF-Files/Distracted-Driving-2009.pdf>
- Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. (2000). *NHTSA Driver Distraction Research: Past, Present, and Future*. Retrieved May 2011, from <http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/233.PDF>
- Ranney, T. (2008). *Driver Distraction: A Review of the Current State-of-Knowledge (DOT HS 810 704)*. Washington, DC: National Highway Traffic Safety Administration.
- Reimer, B. (2010). *Cognitive Task Complexity and the Impact on Drivers' Visual Tunneling*. Washington, DC: Transportation Research Board of the National Academies Publications on Distracted Driving.
- Swallow, D. W., Lindberg, R. M., & Smith-Jackson, T. L. (2003). System Safety Principles and Methods. In H. R. Boomer, *Handbook of Human Systems Integration* (pp. 497-540). Hoboken, NJ: John Wiley and Sons.
- Tijerina, L. (1995). *Final report--program executive summary: heavy vehicle driver workload assessment (DOT HS 808 467)*. Washington, DC: National Highway Traffic Safety Administration.
- U.S. Army Research Laboratory. (2010). Retrieved from Improved Performance Research Integration Tool: www.arl.army.mil/IMPRINT
- U.S. Department of Transportation. (2010). Retrieved May 2010, from <http://www.dot.gov/affairs/2010/dot9910.htm>
- Wickens, C., & Hollands, J. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall.
- Wierwille, L., Tijerina, S., Kiger, T., Rockwell, E., Lauber, A., & Bittner Jr., A. (1992). *Heavy Vehicle Driver Workload Assessment. Task 4: Review of Workload and Related Research (HS-808 467)*. Columbus, OH: Battelle Memorial Institute.
- Wierwille, W. (1993). Visual and manual demands of in-car controls and displays. In B. & Peacock, *Automotive Ergonomics* (pp. 299-320). Washington, DC: Taylor & Francis.
- Wojciechowski, J. (2006). *Validation of a Task Network Human Performance Model of Driving [Electronic Version]*. Retrieved from Virginia Tech: <http://scholar.lib.vt.edu/theses/available/etd-04142006-133043/>
- Wojciechowski, J. (2004). *Validation of Improved Research Integration Tool (IMPRINT) Driving Model for Workload Analysis (ARL-TR-3145)*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.
- Wojciechowski, J., Kogler, T. M., & Lockett, J. F. (2001). *Examining Direct, Teleoperated, and Semi-Autonomous Off-Road Driving From a Human Information Processing Perspective*. Unpublished Manuscript.

3.13 Understanding the Impact of User Frustration Intensities on Task Performance Using the OCC Theory of Emotions

Understanding the Impact of User Frustration Intensities on Task Performance Using the OCC Theory of Emotions

Gloria Washington
The George Washington University
gjwashi@gwmail.gwu.edu

Abstract. Have you ever heard the saying “frustration is written all over your face”? Well this saying is true, but that is not the only place. Frustration is written all over your face . . . and your body. The human body has various means to communicate an emotion without utterance of a single word. The Media Equation says that people interact with computers as if they are human; this includes experiencing frustration. This research measures frustration by monitoring human body-based measures such as heart rate, posture, skin temperature, and respiration. The OCC Theory of Emotions is used to separate frustration into different levels or intensities. The results of this study showed that individual intensities of frustration exist, so that task performance is not degraded. Results from this study can be used by usability testers to model how much frustration is needed before task performance measures start to decrease.

1.0 INTRODUCTION

Interruptions and task-blocking events seem to be common place in human-computer interactions. These task inhibitors contribute to user frustration and are frequent in most users’ interactions with computers. Researchers in human-computer interaction have studied extensively ways to reduce the amount of user frustration experienced by humans; however this emotion continues to occur and varies according to the individual. Since each person is unique, perhaps it is necessary for us to try to understand the relationship between user frustration and its impact on task performance. In this research, the OCC Theory of Emotions is used to separate frustration into its varying amounts, levels, or intensities to understand the unique relationship between each level and user productivity measures.

2.0 BODY

2.1. The Media Equation

The Media Equation theorizes that humans treat new media such as cell phones, tablets, laptops, and video games as if they are interacting with real persons. This display of emotion includes positive and negative emotions. Positive emotions displayed after completing a very important task, signal that activity toward a goal can end and resources can be reallocated for other tasks. However, negative emotions produced after being interrupted from completing a very important task cause an opposite response and motivate users to engage in actions to set things right or prevent unpleasant events from occurring. *User frustration* is one such negative emotion that can occur in human-computer interactions. It is the frustration that occurs when a human is interacting with a computer system. It is experienced by users of any electronic device;

however it is most often associated with desktops, laptops, tablets, and any other portable media.

2.2 Affective Computing

Affective computing is interested in giving computers the ability to interpret human emotions from human body-based measures such as heart rate, skin temperature and respiration measures. User frustration has been measured physiological indicators such as increased heart rate, skin temperature, and respiration. The human body has various ways to communicate user frustration. Some of these ways include increasing the amount of blood pumped to the heart, thereby increasing the temperature of the skin, and increasing the amount of oxygen taken in by the lungs.

Most research in human-computer interaction and affective computing has focused on ways to reduce the amount of user frustration experienced by a human. However, examining the relationship between the amounts of user frustration and task performance measures has not been studied. This paper outlines the use of physiological indicators of frustration, along with the OCC Theory of emotions to model how much user frustration is needed before task performance measures start to decrease. The relationship between user frustration and task performance is also explored in this research. Specifically, the OCC Theory of Emotions was used to separate frustration into various intensities or amounts to understand the individual impact of each level of frustration on task performance measures.

2.3 OCC Theory of Emotions

The OCC Theory was developed in 1988 by Ortony, Clore, and Collins as a cognitive structure to understanding emotions and how they are displayed. Under this theory, there are 22 emotion types and each emotion type has different factors affecting the

intensity of an emotion. Factors that tend to increase emotion intensity often increase the potential for other emotions to occur in that emotion type.

OCC was created for use in artificial intelligence research for emotion synthesis. A key premise in this theory, is emotions have not occurred until they have reached and surpassed a set threshold, Figure 2. Also, once emotions have surpassed this threshold, they are activated and taken on various intensities or amounts as calculated in the original OCC Theory.

Another key premise in OCC Theory is that people perceive the world to be events, agents, and objects and emotion are tightly coupled together. Emotions occur due to consequences of events, actions of agents, or aspects of objects. For instance, a student may have a deadline to submit his/her homework assignment and his/her computer crashes. Initially, the student may feel extreme anger towards his/her computer. However, as the consequences of the event unfold, he/she may begin to feel fear over missing the submission deadline. In the example given, the computer is the agent, the action of the agent is the computer crashing, and failing the homework assignment is the consequence of the computer crashing.

3.0 METHODS

The objective of this work is to try and understand the relationship between the amount of user frustration experienced and task performance. To study this, participants were asked to interact with an application that included system delays of 500 ms while wearing the BioPac BioHarness. Individuals would first experience a control condition that would account for the user's baseline mood and other factors that could affect the test. The actual experiment part of the study included adding the system delays of 500 ms to communications between the mouse and the keyboard.

Measures such as heart rate, skin temperature, posture, and breathing rate were collected from each individual. Productivity measures were calculated as well including: the number of times a subject skipped a task, consecutive number of typos, number of formatting errors, number of uncompleted task, number of times student did not follow directions, etc. These measures were taken during the control and experiment part of the study and then used along with intensities to determine what intensities caused task performance to decrease.

OCC Theory was also adapted for the emotion type anger since frustration is not included in the original theory, Figure 1, and assumed the control study without the added frustrating events provided a baseline for the subject's usual behavior. This was also done for task performance measures.

```

if (value > control high)

intensity = value - control high

else if (value < control low)

intensity = control low - value

else intensity = 0;

```

Figure 1. Adaptation of OCC Theory of Emotions.

To understand the relationship between task performance and user frustration, we counted the number intensities where task performance metrics remain unchanged, the number of intensities where task performance decreased, and the number of intensities where task performance increased and compared these measures for each individual in the study.

44 subjects participated in the study, females and 23 males, and two were excluded because their task performance data was missing or incomplete. Bio-signal intensities were calculated for each of the subjects in the study and task performance was used to compute the total number of intensities that decreased, increased, or left task performance unchanged.

```

IF (EMOTION-POTENTIAL) > (EMOTION-THRESHOLD)
THEN
SET (EMOTION-INTENSITY) = (EMOTION-POTENTIAL) -
(EMOTION-THRESHOLD)
ELSE
SET (EMOTION-INTENSITY) = 0;

```

Figure 2. Original OCC Computational Theory of Emotions.

The lab practitioner conducting the study asked informal questions after the study to understand the mood of the user upon finishing. Users were asked to provide their observations about their level of frustration while interacting with the web-page that included system delays. During the study, users were given the option to skip a task and go on to the next tab, if they did not want to complete the task.

This study is in compliance with the George Washington University's Office of Human Subjects Internal Review Board. All applicable documentation

can be found against the IRB number 110732. All human subjects were provided with an approved informed consent form.

4.0 RESULTS

Results from the study showed that on average users were able to complete the task assigned to them without going over the allotted time of 15 minutes. The duration of the study with added delays was faster. This was surprising since the study included system delays that created a barrier to task completion.

The majority of participants did not skip a task. When asked why participants did not skip a task and go on to the next, some subjects said they wanted to follow directions and felt that the task was easy enough without skipping a task. A few users opted to skip a task, but then went on to complete the skipped task before the end of the study. The only participant to skip a task and stay with his decision was the oldest participant, 45 years of age, in the study.

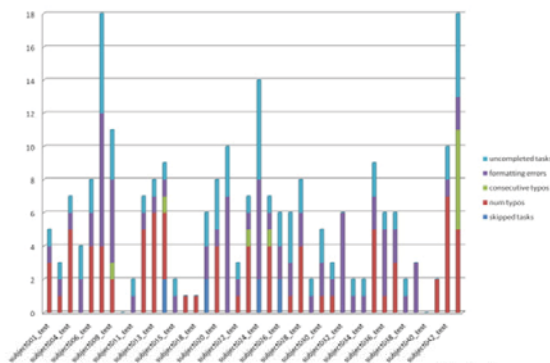


Figure 3. Task performance measures for study with delays.

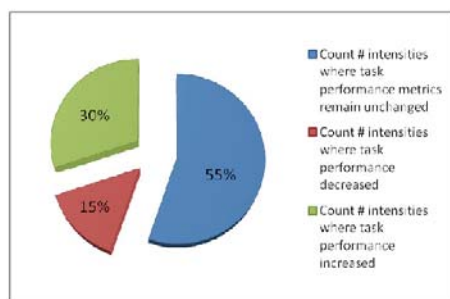


Figure 4. Results for frustration intensities and task performance.

Results from the study showed that users' observations and lab practitioners frustration

observations taken after the experiment were not always congruent with the average intensity calculated by the OCC Theory of Emotions.

5.0 CONCLUSION AND DISCUSSION

Users are more familiar with frustrating incidents than ever before due to the ubiquity of computers. Although developers, testers, and usability engineers have improved the computing interfaces, frustrations still linger. With the everyday use of cell phones, TVs, tablets, mp3 players, etc., users have all become accustomed to certain frustrations. Although frustration occurs as shown in the graphs above, it seems that the normal user, as observed in this study, has adapted themselves to the events. Perhaps, it is necessary to understand and interpret the frustration events and limit their frequency rather than adapting the interface.

From analysis of the graphs above, there seems to exist intensities that allow a user to be productive without affecting task performance. These intensities can better assist usability testers and developers if one can maximize the occurrence of such intensities.

6.0 FUTURE WORK

Future work from this research includes using bio-signal data, task performance measures, and contextual data about what artifacts or widgets a person is interacting with while using an application. Perhaps including contextual information, along with bio-signal and task performance measures could allow developers to gain more insight into what widgets users are encountering the most problems. Also, collecting localized task performance measures along with bio-signal measures taken at longer intervals will help relate these two measures together. These measures, if used during usability testing, could provide a more useful metric in determining how a person really feels about a new application.

An example of this would be in human-in-the-loop usability testing by providing a new user with a system that they have never worked with before. Contextual information such as what widget a person clicked is recorded, along with the number of discrete measurable tasks completed and the time in between each task completed, along with on the fly calculations of user frustration intensities. Since each user is different, perhaps if a certain population of the users are experiencing frustration levels

above a certain threshold set by usability engineers, developers could leverage this information to analyze parts of the user interface that is causing the decrease in productivity measures.

7.0 ACKNOWLEDGMENT

The author of this paper would like to thank Booz Allen Hamilton. This work was funded in part from a grant from the Booz Allen Virginia Center of Excellence.

8.0 REFERENCES

- [1] Anttonen, J and V Surakka. "Emotions and Heart Rate While Sitting on a Chair." Human Factors in Computing Systems, Portland, 2005.
- [2] Ceaparu, I, et al. "Determining Causes and Severity of End-User Frustration." Human Factors in Computing Systems, 2004. 333-356.
- [3] Haag, A, et al. "Emotion Recognition Using Bio-Sensors: First Steps Towards an Automatic System." Lecture Notes in Computer Science in Affective Dialogue Systems (2004).
- [4] Hazlett, R. "Measurement of User Frustration: A Biological Approach." Human Factors in Computing Systems. Ft Lauderdale, 2003. 734-735.
- [5] Klein, J, Y Moon and R W Picard. "This Computer Responds to User Frustration: Theory, Design, and Results." (2001): 119-140.
- [6] McCraty, R, et al. "The Effects of Emotions on Short-term Power Spectrum Analysis of Heart Rate Variability." The American Journal of Cardiology (1995).
- [7] McLaughlin, M, et al. "Recognizing User State from Haptic Data." Cybernetics and Information Technologies. 2004.
- [8] Ortony, A, G L Clore and A Collins. A Cognitive Structure of Emotions. Cambridge: Cambridge University Press, 1988.
- [9] Picard, R W. Affective Computing. Cambridge: MIT Press, 1997.
- [10] Picard, R W and J Klein. "Computers that Recognize and Respond to User Emotion: Theoretical and Practical Implications." MIT Media Lab Tech Report 538. 2003.
- [11] Reeves, B and C I Nass. The Media Equation: How People Treat Computers and New Media Like Real People and Places. New York: Cambridge University Press, 1996.
- [12] Reynolds, C. "The Sensing and Measuring of Frustration with Computers." Master's thesis. 2001.
- [13] Riseberg, J, et al. "Frustrating the User on Purpose: Using Bio-signals in a Pilot Study to Detect the User's Emotional State." Human Factors in Computing Systems. Los Angeles, 1998.
- [14] Scheirer, J, et al. "Frustrating the User on Purpose: A Step Toward Building an Affective Computer." 2001.

3.14 Automatic Facial Expression Recognition and Operator Functional State

Automatic Facial Expression Recognition and Operator Functional State

Nina Blanson
Yale University
nina.blanson@yale.edu

Abstract. The prevalence of human error in safety-critical occupations remains a major challenge to mission success despite increasing automation in control processes. Although various methods have been proposed to prevent incidences of human error, none of these have been developed to employ the detection and regulation of Operator Functional State (OFS), or the optimal condition of the operator while performing a task, in work environments due to drawbacks such as obtrusiveness and impracticality. A video-based system with the ability to infer an individual's emotional state from facial feature patterning mitigates some of the problems associated with other methods of detecting OFS, like obtrusiveness and impracticality in integration with the mission environment. This paper explores the utility of facial expression recognition as a technology for inferring OFS by first expounding on the intricacies of OFS and the scientific background behind emotion and its relationship with an individual's state. Then, descriptions of the feedback loop and the emotion protocols proposed for the facial recognition program are explained. A basic version of the facial expression recognition program uses Haar classifiers and OpenCV libraries to automatically locate key facial landmarks during a live video stream. Various methods of creating facial expression recognition software are reviewed to guide future extensions of the program. The paper concludes with an examination of the steps necessary in the research of emotion and recommendations for the creation of an automatic facial expression recognition program for use in real-time, safety-critical missions.

1.0 Introduction

The increasing mechanization of control processes in the modern world has lessened the gross total workload of the operator and has increased the breadth of tasks that humans can accomplish. However, despite the operator shouldering an objectively smaller workload than before the advent of automation, human error still accounts for the largest percentage of actions leading to the prevention of mission success. When affected by emotional and environmental influences, the operator can falter in the execution of critical tasks. In order to combat the negative consequences of a sub-optimal operator state, it becomes necessary to create a system that both evaluates the operator and influences him or her into embracing an optimal emotional state. Operator Functional State (OFS) defines the current theory of what constitutes an optimal state: "the multidimensional pattern of human psycho-physiological conditions that mediates performance in relation to physiological and psychological costs, [and] results from the synthesis of operator characteristics, current operator conditions, and the operator's interaction with operational requirements." [13] OFS does not place value judgments on a specific aspect of state, but rather looks at its holistic

representation and effect on the operator. The regulation of OFS in operators could help to spot and prevent any human errors from occurring.

Various techniques have been proposed and utilized to measure OFS in real-time, such as electroencephalography (EEG), eye activity, and core temperature measurements. [13] The implementation of these techniques allows for the least noisy sources of data; however, each method only measures a single aspect that affects OFS, and the implementation of the multiple apparatuses measuring these different methods onto a single person would cause considerable discomfort and diminishing returns in respect to the amount of useful data acquired. The most important component of assessing OFS becomes finding a measurement that can take multiple factors into account while extrapolating from a single source of information. This source of information can be found in the human face.

Humans have residual information about their current activities and emotions stored in their face for a variable length of time, and can be charted through tracing muscle movements. Researchers have employed electromyography (EMG) in order to measure muscle activity in the face because of its reliability in detecting facial muscle movement. [9] However, the method

uses electrodes attached to the face, which has several technical and practical drawbacks.[9] In the setting of a flight deck, for example, the inconvenience of attached electrodes overshadows the benefits. A potential non-obtrusive system comes in the form of monitoring an operator through live-video streams, and by using facial expression as a measure of state. This paper overviews the science behind facial expressions, their link to emotions, and describes stress, fatigue, and complacency: three psycho-physiological states that promote sub-optimal OFS. Next follows a description of the feedback loop video system, along with the suggested protocols and models for the regulation of an operator's state in the case of stress, fatigue, and complacency. Then, a description of the basic program created by the author, and an explanation of the computer algorithms and statistical models available to create a facial expression recognition program shows the extent of the possibilities in computer vision.

2.0 Emotion and OFS

Darwin's work, *The Expressions of Emotion in Man and Animals*, postulates that human emotions, and subsequently the physical generation of emotion through facial expressions, are "evolved and adaptive." Darwin proposes that human emotion and its expression could not be the product of societal constructs, but represent the culmination of millions of years of evolution selected to benefit the species.[8] As social creatures even before the advent of human speech, ancestral humans could communicate their emotions and intentions to others through their facial expressions.[8] Despite the formation of language, humans keep their ability to show their emotional state through the face. Ekman and his colleagues later revived interest in Darwin's theory of universality of emotion by continuing cross-cultural studies and noting the similarity of emotional expressions throughout the world. Ekman and his colleagues agreed with Darwin and proclaim

that facial expressions, and the emotions displayed by each facial expression, are universal.[10] In order to create a scientific and methodological manner of determining a facial expression, Ekman also created the Facial Action Coding System (FACS), which uses facial muscle movements, defined as Action Units (AUs), to discern emotion. Vetted by the scientific community, FACS presents a foundation for all measures of facial expression and continues to be a measuring scale for human emotion. FACS allows the scientific examination of facial muscles in facial expression analysis, which can be used to measure OFS after examining the effects of emotion.

Emotion has a continual effect on an operator's perception of a task and on his or her OFS. Studies have shown that emotional arousal has an effect on the general reactivity and decision-making abilities of an individual, despite the emotion's valence.[16] The impact of emotion on cognitive tasks elevates it to an aspect of state to be examined. The face allows a unique combination of both physiological and psychological changes in the body. For example, the perception of happiness in an individual connotes an emotional recognition fueled by the physiological conformation of different muscles in the face. The universality of facial expressions also validates the creation of a facial recognition system focused on the determination of sub-optimal OFS since it would not have to be calibrated for each individual operator. The different conformation of muscles can be measured through the FACS, which already decoded the different potential muscle movements in the face. Although different states of sub-optimal OFS exist, the following three incorporate a large spectrum of the challenges faced by operators and will be concentrated.

2.1 Stress

The Lazarus definition of stress, "the process of appraising [negative or cumbersome] events, of assessing one's

potential to control or cope with the event, and continuing reappraisal as new information becomes available,” defines stress as a process for dealing with harmful, emotional events rather than a strict emotional state in itself.[14] Stress does not constitute an emotion, but rather combines an amalgam of emotions in reaction to a situation. Stress can arise from a time-restricted and work-intensive environment, like those found in safety-critical missions, and has both psychological and physiological effects on the body that leads to a sub-optimal OFS.

In terms of the psychological effects of stress, the brain attempts to mount multiple defensive measures, such as a reflexive completion of tasks without cognitive input, in order to reduce stress.[14] However, following the failure or lack of initialization of these defensive measures, the mind suffers from a reduced ability to process the environment due to the high workload.

Subsequently, stress affects an individual's physical state as well. Selye's General Adaptation Syndrome (GAS) represents the process in which stress manifests in the physiology of an organism. GAS defines the “progressive responses to prolonged stress in which an organism mobilizes for action and compensates for stress,” and its component, Alarm, Resistance, and Exhaustion (ARE), shown in Fig. 1, delineates the stages of the response.[14] The GAS shows the general wearing down of an organism in the face of stress. Ultimately, when stress fails to be mitigated, the organism reaches exhaustion and becomes unable to respond

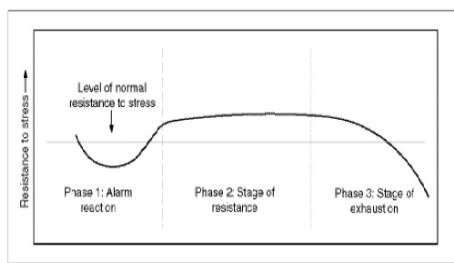


Figure 1. Alarm, Resistance, Exhaustion (ARE).[14] Figure shows the three stages of Selye's General Adaptation Syndrome (GAS).

appropriately to the environment. The culmination of stress, psychologically and physiologically, creates an error-prone environment for the operator due to his or her inability to cope with high-load tasks.

2.2 Fatigue

Fatigue colloquially describes “feelings of fatigue, exhaustion, and to problems of attention and motivation.”[5] However, the amorphous boundaries of this definition of fatigue encompass a multitude of other emotional states like drowsiness and intense hunger. In the context of work-related fatigue, fatigue defines a “normal, healthy response to an oncoming depletion of resources, caused by the execution of physical and mental tasks,” which highlights the fact that humans are not meant to concentrate on a task for an extended period time without regard for their overall state of being.[5] Work-related fatigue does not include the effects of outside influence like lack of sleep, although these factors can be confounded and mistaken as fatigue. Fatigue occurs in a work environment due to extended interaction with a task without any breaks or change of requirements in the task. Fatigue affects the operator's psycho-physiological state by inducing a state of tiredness usually followed by a decrease in cognitive function.

The effects of fatigue leads to the operator being unable to completely process his or her environment. The monotony of the task causes the operator to revert to more energy-saving strategies. For example, in response to the decrease in awareness, the operator either continues the task at his or her current pace and makes mistakes, or slows down the process in an attempt to reduce the mistakes.[5] Essentially, the operator has to take the time to evaluate different work strategies to combat the side effects of fatigue.

2.3 Complacency

Complacency, in the terms of safety-critical missions, defines the tendency for

operators to blindly trust automation without monitoring, and often renders the operator unable to quickly respond to any malfunctions.[12] Complacency also results when the operator works in a highly reliable and automated environment, and the operator's sole task is to stay vigilant in case of occasional error.[15] Although automation has increased the capability of an operator in a multitude of environments, it has also caused an unbalanced mental workload and reduced situational awareness in operators.[12][15] Complacency differs from boredom due to the implicit trust in the automation; complacency lulls the operator into a false sense of security due to trust. While an operator may be bored because of the absence of a task, a complacent operator can also be bored due to the trust given to the automation. A complacent operator appropriates a small cognitive investment in his or her environment, and can lead to overlooking any number of mistakes in the automation.

Complacency does not necessarily increase the number of errors committed by the operator, but rather causes the operator to ignore potentially dangerous situations due to a lack of cognitive awareness. Complacency leads to a reduced reaction time if any errors occur, or an inability to spot any errors at all. If an error occurs while the operator is complacent, the operator could be sent into a panic while attempting to discover and fix the problem occurring in the task simultaneously.

3.0 Video Systems and Facial Expression Recognition

3.1 The Feedback Loop

The use of a feedback loop allows information gathered from the user to be analyzed in real-time and sent back to inform him or her to modify or continue his or her current actions. The proposed feedback loop, as shown in Fig. 2, notifies the user about their state based on the data analyzed from the facial expression

Figure 2. The Feedback Loop. The feedback loop utilized in the potential facial recognition system in order to aid pilots and other critical-safety task workers.

recognition program. The program then sends the product of the data analysis back to the user in the form of useful information regarding their current state. For example, when the subject displays a sub-optimal state of awareness, the feedback loops allows this information to be sent back to the user. Users can then modify their state based on their newly acquired knowledge, or, if they lack the ability due to extreme conditions, allow other protocols to be put into place. The feedback sent to the operator would depend on the emotion he or she displayed, and would be designed to combat that specific sub-optimal state. The different protocols that would be assigned to the three main sub-optimal OFS, and the make-up of the physical systems in the feedback loop, are explained further.

3.2 Suggested Protocols for Sub-Optimal States

3.2.1 Stress Protocol Description

The weight of stress on an operator leads to a variety of negative emotional responses such as fear, anger, disgust, distress, and helplessness.[4][11] Although a distinct stress progression does not project itself directly onto the face, the stress can be traced through the shifts between negative emotions on the face in response to high-load work. The program would monitor the operator for any sequence of negative facial expressions over a period of time. Current research shows that the detection of stress through

live-video streams is both possible and improving and could be an integral part of assessing OFS.[4]

The program would constantly monitor for an emotional reaction to stressors, like a sudden change in the environment, and note the duration of the emotional response. If stress episodes are infrequent and short in duration, the program will not report those results to the operator. However, once the length or frequency of the episodes increase, then the program would report the findings to the operator through voice-command or through a notice in one of the display monitors. The program then prompts the operator to commence a relaxation technique, such as deep breathing or voluntary and directed muscle contraction. The implementation of the stress protocol hopes to reduce stress as it begins, rather than allowing the operator to reach maximum stress levels at any point. However, if the levels of stress are deemed to be overwhelming the operator, then the program will suggest replacing the operator if possible.

3.2.2 Fatigue Protocol Description

Although the measurement of fatigue should not include the symptoms of sleep loss or an out-of-sync circadian rhythm, the outward manifestation of these states are similar.[5] Therefore, for the purposes of the fatigue protocol, all states that manifest similarly to fatigue will be considered as such. However, while fatigue can be alleviated through the implementation of short breaks or modifying the task, the curing of symptoms caused by sleep loss require more intensive recovery. In light of the similarities between the states, the program will also pay attention to the duration that the emotion outwardly affects the operator.

Although the current literature does not describe the facial expressions associated with fatigue, the proposed program looks for characteristics like heavy-lidded eyes combined with a neutral expression and possibly the inability to

detect the operator's face for increments of time due to head-nodding or complete eye closure. The program monitors the operator for their level of fatigue throughout the task. The goal of the program is to squash any signs of fatigue before they affect the operator in a fundamental way and reduces work efficiency. If the level of fatigue passes a level to be determined, the program would inform the operator and instruct them to take a break, focus on another task, or instruct the task-manager to appropriate another task to the operator.

3.2.3 Complacency Protocol Description

Complacency generally manifests itself in ways similar to boredom. In terms of OFS, complacency usually occurs from the under-load experienced due to automation. Research shows that the most effective way to place the operator back into experiencing the appropriate level of load comes in the form of adaptive automation.[12] Adaptive automation refers to the ability to reallocate the level of automation from the operator to the environment depending on the operator's state.[12]

The current literature does not describe the facial expressions associated with complacency, however, for the purposes of the program, complacency will be discerned through the formation of a neutral face, combined with the appearance of cues like constant pupil movement away from potential locations of display monitors. If the operator appears complacent, then the program will alert the adaptive automation system to return some function back to the operator. The transfer of work back to the operator will allow them to regain focus and invest some cognitive awareness into his or her environment.

3.3 The Physical Components

3.3.1 Webcam

Since their introduction, webcams have improved in both their popularity and

their capabilities; however, they still suffer from a multitude of technological shortcomings, such as the maximum number of frames per second and megapixels. The webcam also experiences some practical limitations, like the inability to function well in poorly lit environments.[9] However, despite these setbacks, webcams allow an inexpensive alternative to expensive video equipment; webcams are built to stream video live onto another source, while more traditional video equipment solely stores video for later use. Webcams also have the advantage of size, and the introduction of a webcam into the work environment will be almost unnoticeable due to their small size.

3.3.2 Thermal Camera

Preliminary research shows that thermal cameras present another non-obtrusive way to detect facial muscle movement.⁹ By analyzing the muscles delineated in the FACS, the researchers found that the production of AUs creates an increase in the temperature in certain areas of the face while decreasing other areas, and shows a promising method of evaluating facial muscle movement.[9] Programs that rely on webcams suffer from the camera's inability to register images correctly when the subjects are under poor lighting. Thermal cameras, however, do not require any light at all; the images form through the analysis of heat emanating from a surface. A drawback of thermal cameras comes from the camera's inability to see through other objects. For example, if an operator had to wear goggles, the thermal camera would only reflect the heat of the goggles rather than any of the eye muscles due to the goggles occluding the eyes. The occlusion occurs for all objects, regardless of transparency, so any accessories that cover the face will have to either be removed or redesigned to prevent facial occlusion.

The combination of both standard webcams and thermal cameras in the system would allow for both the continual

monitoring of the operator despite lighting conditions and the ability to corroborate facial muscle movements between the webcam and the thermal camera when the operator sits under good lighting. The utilization of both methods would help to cover some of the drawbacks of either method and create a more balanced system.

4.0 Computer Algorithms for Deciphering Facial Expressions

The current iteration of the automatic facial recognition program, created in Visual C++, utilizes Open Computer Vision (OpenCV) libraries and Haar cascades, and tracks a human face in a live video-feed. After identifying a human face, the program then subsequently identifies some key facial landmarks: the eyes, pupils, nose, mouth, and the corners of the mouth. Some of the limitations of the program, both intentional and unintentional, are as follows: the program currently only tracks one individual at a time, has a slight delay due to the application of the facial landmark identification boxes onto a live-stream video, does not include thermal camera integration, and can only identify facial features deviating 30 degrees either to the left or right from facing the camera directly.

The program also lacks a key component: facial expression recognition. The current state of computer vision allows for the utilization of a multitude of algorithms to create a program that automatically recognizes facial expression. The most prominent algorithms seen in recent years are Haar cascades, Active Shape Models, Active Appearance Models, and the Piecewise Bézier Volume Deformation Model.

4.1 Haar Cascades

Haar cascades, described by Viola and Jones [2001], describe a method to quickly identify images using feature-detection. Haar cascades rely on Haar-like features that detect the change in contrast

values between adjacent rectangular groups of pixels, rather than the intensity values of a pixel itself.[19] Although the use of Haar-like features reduces computing time drastically, comparing every feature in an image in order to find an object creates unnecessary work.[19] The cascading of classifiers allows the further reduction of necessary computing power. Cascading refers to the utilization of a degenerate decision tree that processes images through a series of comparison tests, and rejecting the image if it fails any one of the tests.[18] For example, if the part of the image being compared matches the feature in the first sub-window but does not match the second, then the cascade rejects that portion of the image and moves to the next feature. The degenerate decision tree allows for the comparison of all pixels without an unnecessarily complete classifier comparison, and results in the rejection of most negative images and the approval of most positive images.

In order for a Haar classifier cascade to identify a particular object, the algorithm has to be trained. The training of the algorithm requires a large library of both positive and negative images. The positive images contain the object that the user wants the algorithm to detect, while the negative images consist of random images without the object.[18] The robustness of the algorithm has a positive relationship with the number of images used for training; however, this positive relationship occurs due to the larger probability of variety in larger samples. In the case of facial feature detection, the sample images would have to span across ages, genders, and races.

For facial expression detection, however, Haar cascades create some difficulty. Facial expressions, while universal, can also be repressed or faked to some degree. While noses and eyes vary solely in shape and size, facial expressions vary in the totality of their expressions, from the arch of the brow to the movement of the mouth, and the differences between two different emotions could be too subtle for the Haar cascade to differentiate between.

Instead of whole-face facial expression recognition, it could be possible to track more emotive landmarks of the face, like the eyebrows, and then create a program that would note the locations of facial features in relation to the proportions of the face, and then process that information for emotional detection. However, that process would be tedious and probably inaccurate for the subtleties of emotional expression.

4.2 Active Shape Models (ASM)

Cootes and Taylor's [1995] ASMs illustrate a trained statistical modeling algorithm that utilizes Point Distribution Models (PDMs) to rapidly locate the desired structure.[3] PDMs describe a training set of images dotted with landmark points. The training sets have multiple instances of the same object, but with some natural variation. The mean position of the landmark points in relation to the object, and then the points that show how each object in the training set deviates from the calculated mean, are determined.[3] After the establishment of the PDMs, the ASM performs an iterative search to locate the shapes of the structures in different images.[3] The ASMs result in the ability to fit the model to the image after a finite number of iterations.

The human face, although variable, generally follows a model format with the same number of facial features. The general shapes of the facial features are also the same, although the size of these features in relation to each other, and in general, can change. ASMs then provide a non-rigid modeling approach to identify and conform to different facial expressions. The measured facial expressions would be trained by different sets of PDMs and then the conformation of certain ASMs would denote that expression.

However, the use of ASMs in order to describe human emotion runs into similar problems as Haars cascades – although ASMs require less training, they only identify models defined in their training sets. Therefore, a PDM would have to be created

for each emotion. Even though ASMs are more forgiving than Haars cascades when conforming to images, the complexities of an emotion would have too much deviation from the mean. For example, the manifestation of a smile can range from the open mouth smile to the Duchenne smile. Each variation of the smile still constitutes a smile; however, the difference between a closed-mouth smile and an open-mouth smile, combined with the appearance of teeth, only constitute a portion of the variation.

4.3 Active Appearance Models (AAM)

AAM, also created by Cootes and Taylor [2001], shows a statistical modeling method similar to ASM, but diverges from tracking a specific object and moves toward matching a class of deformable objects.[2] AAM works similarly to ASM by requiring a training set marked with points. An image in the training set, after being marked, has its points aligned; the subsequent wrapping of the images around the mean of the training set then creates a “shape-free patch,” which lacks any texture.[2] The process is repeated for each image in the set, training the model to conform to a variety of shapes of a similar class. The “shape-free patch” is raster scanned into a texture vector and then applied with Eigen-analysis to build a texture model.[2] Lastly, the correlations between shape and texture generate a combination of the shape and the texture to create a combined appearance model.[2]

The ability to follow deformable objects, such as a face with any expression, bestows the ability to identify the expression without specific training, allowing AAMs to surpass ASMs in facial expression recognition. AAM works slightly slower than ASM, but the added robustness and the reduction in different training sets for different emotions overrules any advantages of ASM in terms of identifying facial expression.[2] AAM, however, does require a large set of images for training, and the availability of large facial databases are limited.[17] Until more high-quality

databases are established, AAMs will not be able to attain the robustness capable of the model.

4.4 Piecewise Bézier Volume Deformation Model (PBVDM)

The PBVDM utilizes the Connected Linear Deformation Method (CDLM), from the work of Tao and Huang [2003] and defines a model consisting of multiple deformable 2D patches whose movements are determined by a set of different vibration modes.[7] Vibration modes refer to the different displacement vectors that determine the deformation of a particular patch. These patches are connected together with hinges that work to maintain the spatial relationships between the patches and prevent any drifting. CDLM can therefore be used to track and model the movements of facial muscles, which would be represented as different vibrations of the patches.[7]

The PVDM creates Bézier volumes across specified areas of the face, with a section of the surface model, or facial mesh, in between the top and bottom volume layers. The movement of the Bézier volumes deforms the surface model, and in the case of facial mesh, shows the movement of facial muscles in that region. PBVDM can utilize the AUs of the FACS to define the available facial movements. Therefore, the PBVDM allows a more specific analysis of desired facial muscle movements in real-time.[6]

The use of PVDM in facial expression recognition allows a more scientific approach. The other models require training sets to tell the model what constitutes an emotion, and then the models can only identify those trained emotions. PVDM, on the other hand, allows an analysis of muscle movements through the use of individual volume patches without training sets. PBVDM can be made to be analogous with a FACS analysis of a face. Current PBVDM facial meshes do not cover the complexity of facial muscle movement covered by FACS, or the ability to decipher

emotions as effectively as a human can, but the accuracy of the algorithms is improving.

5.0 Discussion

Currently, the research surrounding the emotional states of boredom and fatigue are limited in both their breadth and depth. More research into the states and facial expressions associated with fatigue and complacency would be helpful in creating algorithms to define those emotional states. A potential method of determining the facial expressions associated with fatigue and complacency would be to conduct a study where individuals would be exposed to situations that lead to those emotional states, and then facial expression would be analyzed to search for commonalities, similar to the current studies on stress.[4]

The current iteration of the program developed by the author allows for the detection of some facial landmarks, which represents the first step in automatic facial expression recognition. The program could be extended even further to recognize other landmarks like the eyebrows. In the future, the actual creation of an automatic facial expression recognition program for use in human factor's experiments will help in the tracking of sub-optimal states of awareness in operators. The facial expression recognition program should be created through the use of either AAMs or PBVDM due to their robustness and ability to reliably deform to the human face. Unless more facial databases come into fruition, the use of PBVDM would be optimal due to its ability to deform to the human face without the need for high quality training sets.

After the creation of a proper facial expression recognition program, a complete study to evaluate the emotion protocols and their effect on the frequency of errors during flight would help to determine the efficacy of the system.

6.0 Conclusion

The study of OFS requires the consideration of a large number of variables in order to find the optimal OFS. However, the use of automatic facial expression recognition could help to reduce the number of accidents due to human error by helping to reduce sub-optimal states of OFS before they overtake the operator. The use of video systems in lieu of more obtrusive methods of determining state allows the practical implementation of state-measuring devices into the operator's environment. The implementation of video-systems in a work environment would help to increase the safety experienced by both operators and their charges, and lead to a safer world.

References

- ¹Azcarate, A., Hageloh, F., van de Sande, K., Valenti, R., "Automatic Facial Emotion Recognition," Universiteit van Amsterdam, Netherlands, 2005.
- ²Cootes, T. F., Edwards, G. J., Taylor, C. J., "Active Appearance Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, 2001, pp. 681-685.
- ³Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J., "Active Shape Models – Their Training and Application," *Computer Vision and Image Understanding*, Vol. 61, No. 1, 1995, pp. 38-59.
- ⁴Dinges, D. F. et al., "Optical Computer Recognition of Facial Expressions Associated with Stress Induced by Performance Demands," *Aviation, Space, and Environmental Medicine*, Vol. 76, No. 6, Section II, 2005, pp. B172 – B182.
- ⁵Gaillard, A. W.K., "Fatigue Assessment and Performance Protection," *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*, Series I: Life and Behavioral Sciences, Vol. 355, edited by Hockey, G. R., Gaillard A. W.K., Burov O., IOS Press, Amsterdam, Netherlands, 2003, pp. 24-35.
- ⁶Hai, T., Huang, T.S., "A Piecewise Bézier Volume Deformation Model and its Applications in Facial Motion Capture," *Advances in Image Processing and Understanding: A Festschrift for Thomas S. Huang*, edited by A. C. Bovik, C. W. Chen, and D. B. Goldgof, World Scientific Publish Co. Pte. Ltd., 2003, pp. 39-56.
- ⁷Hai, T., Huang, T.S., "Connected Vibrations: A Modal Analysis Approach for Non-Rigid Motion Tracking," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998, pp. 735-740.
- ⁸Hess, U., Thibault, P., "Darwin and Emotion Expression," *American Psychologist*, Vol. 64, No. 2, 2009, pp. 120-128.
- ⁹Jarlier, S. et al., "Thermal Analysis of Facial Muscles Contractions," *IEEE Transactions on Effective Computing*, Vol. 2, No. 1, 2011, pp. 2-9.
- ¹⁰Keltner D., Ekman P., "Facial Expressions of Emotion," *Handbook of Emotions*, 2nd ed., edited by M. Lewis and J. Haviland-Jones, Guilford Publications, Inc, New York, 2000, pp. 236-245
- ¹¹Lerner, J. S., Dahl, R. E., Hariri, A. R., Taylor, S. E., "Facial Expressions of Emotion Reveal Neuroendocrine and Cardiovascular Stress Responses," *Biological Psychiatry*, Vol. 61, No. 2, 2007, pp. 253-260.
- ¹²Parasuraman, R., "Adaptive Automation Matched to Human Mental Workload," *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*, Series I: Life and Behavioral Sciences, Vol. 355, edited by Hockey, G. R., Gaillard A. W.K., Burov O., IOS Press, Amsterdam, Netherlands, 2003, pp. 177-193.
- ¹³RTO HFM Task Group HFM-056/TG-008, "Operator Functional State Assessment," RTO Technical Report TR-HFM-104, Kiev, Ukraine, Sept. 2003.
- ¹⁴Parker, M. M., Ettinger, R. H., "Emotion and Stress," *Understanding Psychology*, 3rd ed., BVT Publishing, Redding, California, 2010, pp. 377-373.
- ¹⁵Pinzel, L. J., DeVries, H., Freedman, F. G., Mikulka, P., "Examination of Automation-Induced Complacency and Individual Difference Variates," NASA/TM-2001/211413, 2001.
- ¹⁶Raddatz, K. R., Werth, A., Tran, T. Q., "The Influence of Emotional State and Pictorial Cues on Perceptual Judgments," *Human Factors and Ergonomic Society's Annual Meeting Proceedings*, Vol. 51, Student Forum, 2007, pp. 1496-1500.
- ¹⁷Ratliff, M. S., Patterson, E., "Emotion Recognition Using Facial Expressions with Active Appearance Models," *Third IASTED International Conference on Human Computer Interaction*, Understanding Users, 2008, pp. 138-143.
- ¹⁸Viola, P., Jones, M., "Rapid Object Detection Using a Boosted Cascade of Simple Features," *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, Kauai, Hawaii, 2001, pp. 511-518.
- ¹⁹Wilson, P. I., Fernandez, J., "Facial Feature Detection Using Haar Classifiers," *Journal of Computing Sciences in Colleges*, Vol. 21, No. 4, 2006, pp. 127-133.

Acknowledgements

The author would like to thank her mentors, Dr. Alan Pope and Chad Stephens, for helping her throughout the project, the coordinators of NASA's Undergraduate Student Research Program (USRP), and NASA.

3.15 A Human View Model for Socio-Technical Interactions

A Human View Model for Socio-Technical Interactions

Holly A. H. Handley, PhD, PE & Andreas Tolk, PhD
Old Dominion University, Norfolk, VA 23529
hhandley@odu.edu

Abstract. The Human View was developed as an additional architectural viewpoint to focus on the human part of a system. The Human View can be used to collect and organize data in order to understand how human operators interact and impact the other elements of a system. This framework can also be used to develop a model to describe how humans interact with each other in network enabled systems. These socio-technical interactions form the foundation of the emerging area of Human Interoperability. Human Interoperability strives to understand the relationships required between human operators that impact collaboration across networked environments, including the effect of belonging to different organizations. By applying organizational relationship concepts from network theory to the Human View elements, and aligning these relationships with a model developed to identify layers of coalition interoperability, the conditions for different levels for Human Interoperability for network enabled systems can be identified. These requirements can then be captured in the Human View products to improve the overall network enabled system.

1.0 INTRODUCTION

Interoperability is the ability of systems to provide services to and accept services from other systems in order to enable them to operate effectively together [1]. Human Interoperability strives to understand the relationships between human operators that impact collaboration among individuals, teams and organizations across technology environments [2]. Human Interoperability is especially critical under conditions where partnerships and teams may need to be formed rapidly, where technological and organizational compatibilities can present constraints, or closely coupled team interactions have to be achieved across remote locations. By understanding and incorporating Human Interoperability requirements into system design, the resulting networked enabled system can more effectively support timely responses to events. Solutions must include process, organization, and people: "People are included as part of the definition of a system, but their role in that system is generally poorly specified, and the focus of the [system] engineering effort is on the technology components" [3].

The Human View architecture represents a methodology to collect and organize human system data for a network enabled system. It was developed to augment existing architectural frameworks with additional information relevant to the human system requirements in net-centric environments [4]. Additionally, research in the network analysis domain combined knowledge management, operations research and social network techniques to create a "Meta-Matrix." The

Meta-Matrix characterizes relationships between the different organizational elements of people, knowledge, resources, tasks, and organizations [5]. By applying these elemental constructs to the Human View, the relationships between the Human View elements can be defined in terms of network organizational analysis. The network definitions from this combined framework can then be mapped to the Layers of Interoperability defined for network enabled systems [6]. This reference model extends traditional technical interoperability models to include organizational aspects important in for coalition interoperability. This methodology is shown in Figure 1.

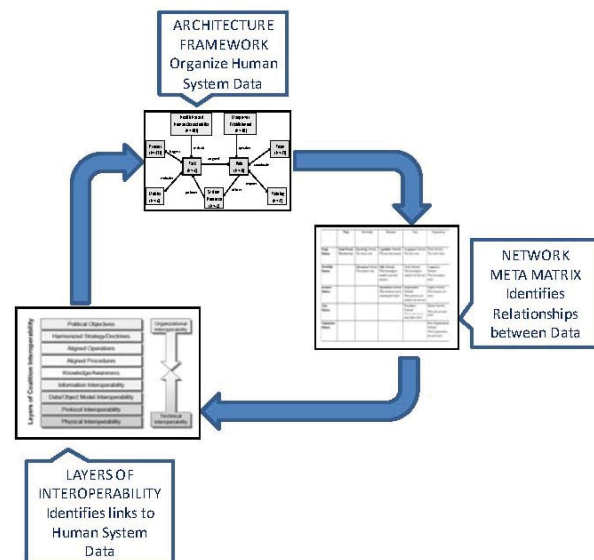


Figure 1. Research Methodology

The composite model resulting from the mapping of the Human View to the Meta-Matrix, and then the Layers of Interoperability can be used to define Human Interoperability goals for a system. The required data that supports these goals can be identified through the model and captured in the Human View products. This information can then be used to inform the system architecture of its strengths and weaknesses with regard to socio-technical design.

The next section will describe each of these models individually, and then explain the methodology used to create the composite model. An example of its use to map the goals for an interagency response to a crisis situation is described, as well as the resulting requirements for the Human View products. The conclusion reiterates the need to continue to define Human Interoperability goals for network enabled systems.

2.0 BACKGROUND

One of the goals driving the development of the original Human View was to create an integrated set of products that could be used to inform and influence system design, especially for network enabled systems. The Human View was designed to organize human information in order to provide a comprehensive representation of human capabilities [4]; the "products" that capture the human information are listed in Table 1. In addition, a representation of the relationships between the products was devised in order to provide a model of the data dependencies; see Figure 2. From this representation, a simulation model can be created to evaluate the impact of the human on the system performance, and to compare alternative human-system configurations [7]. The comprehensive Human View and accompanying simulation model provides an understanding of the human role in the system and supplies a basis for stakeholder decisions by linking the engineering community to the manpower, personnel, training, and human factors communities.

Table 1. Human View Architecture Products [4]

Human View Product	Description
HV-A Concept	High-level representation of the human component(s) in the system
HV-B Constraints	Repository for different classes of human limitations
HV-C Tasks	Describes the human-specific activities
HV-D Roles	The job functions defined for the humans interacting with the system
HV-E Human Network	The human-to-human communication patterns that occur in teams
HV-F Training	Accounting of training requirements, strategy, and implementation
HV-G Metrics	Repository for human-related values, priorities, and performance criteria
HV-H Dynamics	The feeder data and scheme for a simulation of humans in the system

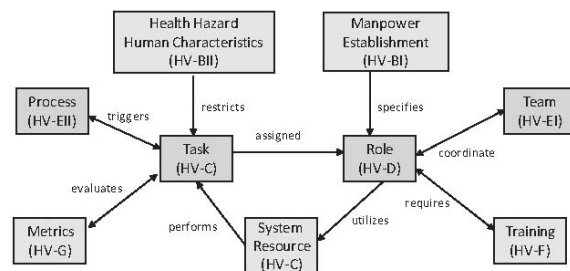


Figure 2. Human View Modeled Dependencies

Network theory has explored the relationships within organizations and had characterized organizational architectures in terms of three domain elements: individuals, tasks, and resources [8]. This was later expanded to include the elements of knowledge and organizations [5]; the resulting extended matrix depicts a set of ten network relationships that connect these elements, as shown in Table 2. This "Meta-Matrix" provides a characterization of the possible networks among the organizational elements using relational primitives and formalizes the dependencies among organizational elements [5]. By applying this matrix to the Human View framework, the Meta-Matrix can provide a way to represent the human system data relationships based on network theory.

Table 2. Extended Meta-Matrix from Network Theory [5]

	People	Knowledge	Resources	Tasks	Organizations
People Relation	Social Network: Who knows who (1)	Knowledge Network: Who knows what (2)	Capabilities Network: Who has what resource (3)	Assignment Network: Who does what (4)	Work Network: Who works where (5)
Knowledge Relation		Information Network: What informs what (6)	Skills Network: What knowledge is needed to use what resource (7)	Needs Network: What knowledge is needed to do what task (8)	Competency Network: What knowledge is where (9)
Resources Relation			Substitution Network: What resources can be substituted for which (10)	Requirements Network: What resources are needed to do what task (11)	Capital Network: What resources are where (12)
Tasks Relation				Precedence Network: Which tasks must be done before which (13)	Market Network: What tasks are done where (14)
Organizations Relation					Inter Organizational Network: Which organizations link with which (15)

Interoperability is often viewed as a problem to be solved through technological solutions. However, true interoperability is achieved through a combination of non-material solutions (organization, process, and people), as well as through technology (data, information, and systems). A model was developed to address the interoperability issues raised with the transition of coalitions to networked enabled systems [6]. Networked enabled systems must focus not only on technology concerns, but also people and process components as well, especially systems that cross organizational boundaries. The reference model of Layers of Interoperability, shown in Figure 3, depicts the interplay of technical and organizational interoperability within a networked enabled environment. The lower levels of the model address layers of technical interoperability, such as the ability to collect and distribute data and information, while the top levels focus more on the non-material requirements, i.e., the organizational interoperability layers dealing with the harmonization and coordination of related operations [6]. This representation of the Layers of Interoperability can be mapped to the Meta-Matrix through the network

definitions. This mapping can then be used to identify the Human Interoperability goals of the network enabled system.

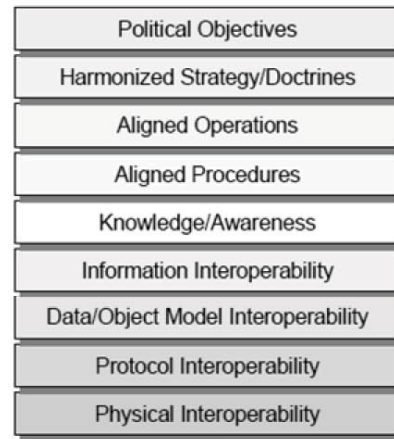


Figure 3. Layers of Interoperability [6]

3.0 METHODOLOGY

The Meta-Matrix relationships are numbered one through fifteen as shown in Table 2; these numbers were used to identify the people, knowledge, resource, tools, and organization relationships in the Human View, depicted in Figure 2. The first step was to align the Human View entities with the Meta Matrix entities. This was

based on the definition of the Human View products, and the result is shown in Table 3. Through the mapping process, several of the Human View entities shown in Figure 2 were "pruned", as these were not relevant to the current mapping; this included the HV-B Constraints and HV-G Metrics elements. The second step was to align the Meta-Matrix relationships with the Human View dependencies; this resulted in the Human View network as shown in Figure 4. In this figure the Human View entity names have been replaced with the Meta Matrix entities, and the dependencies labeled with the network numbers. Additionally, the HV-EI Process has been folded back into the HV-C Task; and indicated with a "self loop" on the Task entity (#13). While this figure validates the mapping of the Human View to Meta-Matrix, the table form of the mapping, similar to Table 2, will be used for the aggregate model.

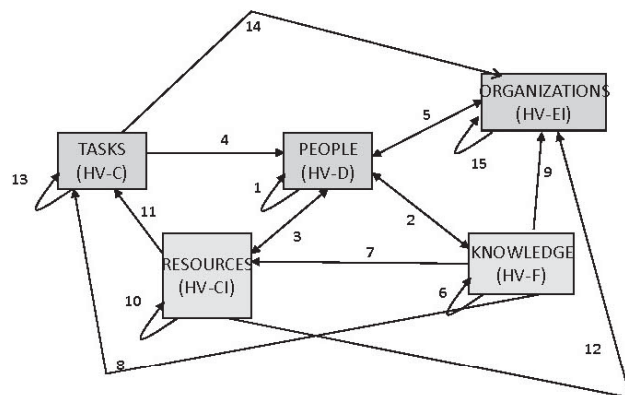


Figure 4. Mapping of Meta Network to Human View Relationships

The next step was to map the five network entities to the Layers of Interoperability model shown in Figure 3. This was done by evaluating the description of each interoperability layer and performing a best match to the entity, as shown in Table 4. This match was then used to identify each of the Layers of Interoperability within the Meta-Matrix. The mapping was limited to the top five layers of the model, which represent the non-materiel interoperability solution.

Table 3. Entity Mapping from Human View to Meta Network

Human View Entity	Meta Matrix Entity
Role (HV-D)	People
Training (HV-F)	Knowledge
System Resource (HV-CI)	Resource
Task (HV-C)	Task
Team (HV-EI)	Organization

Table 4. Layers of Interoperability Description and Entity Assignment

Layer of Interoperability	Description	Meta Matrix Entities
Political Objectives	Do the partners share the same political values?	Organization
Harmonized Strategy	Are the social and cultural backgrounds of the partners aligned?	People
Aligned Operations	Are the decision makers aware of the processes of the partners?	Tasks
Aligned Procedures	Are the tactical requirements supported by knowledge systems?	Resources
Knowledge/Awareness	Are various views on the operation supported?	Knowledge

The result of the composite mapping of the Human View to the Meta Matrix to the Layers of interoperability is shown in Figure 5. The rows represent the organizational elements of the Meta-Matrix with the network relationships labeled in each cell. The columns represent the Human View product that will capture the data to support the relationships identified in each column. The Layers of Interoperability are labeled within the cells, in addition to the network relationships, that support that layer. While there is a one-to-one mapping of the Layer of Interoperability to the organizational entities in the cells along the diagonal, adjacent cells also contribute information to that layer as shown. This composite model will be used in the next section to illustrate how it can drive the Human View

requirements in order to meet Human Interoperability goals.

	Role (HV-C)	Training (HV-F)	System (HV-CI)	Tasks (HV-C)	Team (HV-EI)
People Networks	1 (Social Network) HARMONIZED STRATEGY (HS)	2 (Knowledge Network) KA & HS	3 (Capabilities Network) AP & HS	4 (Assignment Network) AO & HS	5 (Work Network) PO & HS
Knowledge Networks		6 (Information Network) KNOWLEDGE AWARENESS (KA)	7 (Skills Network) AP & KA	8 (Needs Network) AO & KA	9 (Competency Network) PO & KA
Resources Networks			10 (Substitution Network) ALIGNED PROCEDURES (AP)	11 (Requirements Network) AO & AP	12 (Capital Network) PO & AP
Tasks Networks				13 (Precedence Network) ALIGNED OPERATIONS (AO)	14 (Market Network) PO & AO
Organization Networks					15 (Inter-Organizational Network) POLITICAL OBJECTIVES (PO)

Figure 5. Composite Mapping of the Models

4.0 INTERAGENCY EXAMPLE

An interagency response to a crisis situation is an example where increased levels of Human Interoperability, as indicated by the top five Layers of Interoperability shown on Figure 3, can increase the effectiveness of the interagency response. Emergency situations, such as a natural disaster or other crisis situation beyond the scope of local emergency resources, activate the use of an Incident Center (IC) that mobilizes to provide assistance during the event. "Incidents typically include two or more organizations or sub-organizations, each with its own command structure, that respond to the crisis as one unified IC center" [9]. In some cases the interagency response enacted is defined as a Defense Support to Civil Authority (DSCA), indicating military commands are available to provide key support and resources to civilian authorities [10]. The main response phases are described in [9]; however the goals of each phase have been summarized in Table 5. Based on the goal description, it can be mapped to represent a Meta-Matrix relationship, and therefore support one or several interoperability layers, as shown in the second and third columns of the table.

Table 5. Interagency Response Goals aligned with Meta Network and Interoperability

Interagency Response Goals	Meta Matrix	Interoperability Layer
<i>Pre event Planning & Monitoring</i>	<i>Organization & Knowledge Entities</i>	<i>Understand the willingness and abilities of different organizations.</i>
A. Set up a partnership between organizations	Which organizations link with which? (15)	Political Objectives
B. Agree on tasking for situational contingencies	What tasks are done where? (14)	Political Objectives; Aligned Operations
C. Agree on business practices for information sharing	What knowledge is needed to do what task? (8)	Aligned Operations
D. Train staff on procedures from different organizational personnel	Who works where? (5)	Political Objectives Harmonized Strategy
E. Share and synthesize information to identify triggering event	What knowledge is where? (9)	Political Objectives;
<i>Trigger and Progress on Crisis Tasks</i>	<i>Resource & People Entities</i>	<i>Guidance regarding roles, pooling resources, and sharing information.</i>
F. Broaden communication network	What resources are where? (12)	Aligned Operations
G. Establish two-way correspondence between field respondents	Who knows who? (1)	Harmonized Strategy
H. Ability to share field reports and data	Who knows what? (2)	Harmonized Strategy
I. Alert other agencies of the situation progress	Who does what? (4)	Aligned Operations; Harmonized Strategy
J. Re-task resources as necessary	Who has what Resource? (3)	Harmonized Strategy
Conclusion of Crisis Tasking	<i>Task Entity</i>	<i>Complete tasking with depleted resources.</i>
K. Conclude tasking associated with resolving the situation	Which tasks must be done before which? (13)	Aligned Operations

Table 6. Human View Product Requirements base on Interoperability Goals

Human View Product	<i>Role (HV-D): Harmonized Strategy (HS)</i>	<i>Training (HV-F): Knowledge Awareness (KA)</i>	<i>System (HV-CI): Aligned Procedures (AP)</i>	<i>Tasks (HV-C): Aligned Operations (AO)</i>	<i>Team (HV-EI): Political Objectives (PO)</i>
People Networks	G. Establish Two-way Comms (1) HS	H. Ability to Share Field Data (2) KA & HS	J. Re-task Resources (3) AP & HS	I. Alert other Agencies (4) AO & HS	D. Train Different Staff (5) PO & HS
Knowledge Networks		<i>Information Network (6) KA</i>	<i>Skills Network (7) AP & KA</i>	C. Agree on Practices (8) AO & KA	E. Share Information (9) PO & KA
Resources Networks			<i>Substitution Network (10) AP</i>	<i>Requirements Network (11) AO & AP</i>	F. Broaden Communications (12) PO & AP
Tasks Networks				K. Conclude Tasking (13) AO	B. Agree on Tasking (14) PO & AO
Organization Networks					A. Set up Partnership (15) PO
Product Completed	HV-D: "Human Roles Matrix for DSCA Events".	HV-F: "Information Flow for DSCA Events"	<i>No Product Completed</i>	HV-C: "Task to Role Assignment Matrix for DSCA Events"	HV-E: "Team Interaction Matrix for DSCA Events"
Product Description	<i>Identification of people, roles, and contact information to facilitate communication</i>	<i>Information on access and flow of information to facilitate situation awareness.</i>	<i>N/A</i>	<i>Assignment of tasks, and descriptions of processes to facilitate understanding of procedures.</i>	<i>Command and control structure of contributing organizations to facilitate team interactions.</i>

These goals can then be mapped to the composite model to determine the requirements of the Human View products. The Human View products can be either created, or evaluated if they already exist, to determine the necessary data that should be captured in each product to support the Human Interoperability conditions. As shown in Table 6, the mapping of the interagency response goals to the composite model helped define the requirements for each of the Human View products created. The Human View products completed for the DSCA response, as indicated in the row labeled "Product Completed" can be viewed in [9]; the content of each product is described in the "Product Description" row of Table 6.

5.0 CONCLUSION

The Human View was developed as an additional architectural viewpoint to focus on the human part of a system. The Human View can be used to collect and organize social parameters in order to understand the way that humans interact with other elements of the system. The framework can also be used to develop a model of socio-technical interactions and applied to the emerging area of Human Interoperability. Human interoperability strives to improve collaboration among diverse people and teams. Understanding and incorporating human interoperability into system design can assist in the integration and interaction among human operators, improving congruent behaviors for collaborative tasks.

By mapping a matrix of network relationships to the Human View elements and aligning these with a model of coalition interoperability layers, the conditions for different levels for Human Interoperability for network enabled systems can be assessed. An example has shown how, from this mapping, the design variables affecting Human Interoperability can be identified and captured in the Human View products and used to improve network enabled system design. By including Human Interoperability requirements, systems can be designed that facilitate sharing and collaboration through technological environments across teams and organizations.

ACKNOWLEDGEMENTS

The authors would like to thank Nancy Heacox and Chiesha Stevens for sharing information about the DSCA Human View products completed for the Trident Warrior 2006 experiment.

REFERENCES

- [1] Jain, R., Chandrasekaran, A. & Erol, O. (2010). A Systems Integration Framework for Process Analysis and Improvement. *Systems Engineering*, 13(3), 274-289.
- [2] Brown, A. (2010). Human Interoperability and Building Partnership Capacities: Introduction to Human Interoperability, *Human Interoperability and Net-Centric Series*, National Defense University.
- [3] Elm, W., Gualtiere, J., McKenna, B., (2008). Integrating Cognitive Systems Engineering Throughout the Systems Engineering Process, *Journal of Cognitive Engineering and Decision Making*, 2(3).
- [4] Handley, H., & Smillie, R.. (2008). Architecture framework human view: The NATO Approach. *Systems Engineering*, 11(2), 156-164.
- [5] Carley, K. (2002).) Inhibiting Adaptation, *Proceedings of the 2002 CCCRTS*, June, Monterey, CA.
- [6] Tolk, A. (2003), Beyond Technical Interoperability – Introducing a Reference Model for Measures of Merit for Coalition Interoperability, *Proceedings of the 2003 CCCRTS*.
- [7] Handley, H.A.H., & Smillie, R.J. (2010). Human View Dynamics - The NATO approach. *Systems Engineering*, 13(1) p. 72-79.
- [8] Krackhardt, D. & Carley, K. (1998) A PCANS Model of Structure in Organizations, *Proceedings of the 1998 ICCCRTS*, June, Monterey, CA pp. 113-119.
- [9] Stevens, C. & Heacox, N. (2008) Using NATO Human View Products to Improve Defense Support to Civil Authority, *Proceedings of the 2008 ICCCRTS*, June 17-19, Bellevue, WA.
- [10] Dourandish, R., Zumel, N. & Manno, M (2007) Command and Control of the First 72 Hours of Disaster Response, *Proceedings of the 2007 ICCCRTS*, June 19-21, Newport, RI.

3.16 Understanding Group/Party Affiliation Using Social Networks And Agent-Based Modeling

Understanding Group/Party Affiliation Using Social Networks And Agent-Based Modeling

Kenyth Campbell
Old Dominion University
kenyth.campbell@gmail.com

Abstract: The dynamics of group affiliation and group dispersion is a concept that is most often studied in order for political candidates to better understand the most efficient way to conduct their campaigns. While political campaigning in the United States is a very hot topic that most politicians analyze and study, the concept of group/party affiliation presents its own area of study that produces very interesting results. One tool for examining party affiliation on a large scale is agent-based modeling (ABM), a paradigm in the modeling and simulation (M&S) field perfectly suited for aggregating individual behaviors to observe large swaths of a population. For this study agent based modeling was used in order to look at a community of agents and determine what factors can affect the group/party affiliation patterns that are present. In the agent-based model that was used for this experiment many factors were present but two main factors were used to determine the results. The results of this study show that it is possible to use agent-based modeling to explore group/party affiliation and construct a model that can mimic real world events. More importantly, the model in the study allows for the results found in a smaller community to be translated into larger experiments to determine if the results will remain present on a much larger scale.

1. INTRODUCTION

Political party affiliation is a huge part of the American political system, not to mention other political systems around the world. If it was possible to simulate party affiliation dispersion patterns, one could more efficiently promote political candidates. The following research project investigates political party affiliation dispersion patterns in local communities. As a side note, it needs to be mentioned that the original intent for the model changed as research progressed.

The original idea for the model was to look at all the factors that contribute to political party affiliation and essentially input these variables into a Netlogo [12] model. These variables looked at how people in a local community responded to influence from two different groups. The first of these groups are immediate neighbors. For this study neighbor is defined as an individual living in a person's immediate vicinity, i.e. next door or across the street. Also, immediate neighbors are not

chosen by individuals, they are simply the people already in place when a person moves into a local community. The second variable consisted of friends present in an individual's social network. A person's social network consists of friends and family that a person chooses to associate with, and chooses from whom they will receive advice or influence.

The concepts stated above include many complex ideas and attitudes that must be attributed to everyone in the system that is to be modeled. Based on this fact, the agent based modeling (ABM) paradigm made the most sense for this research project. ABM allows for many agents to be created in a single system. Since the research question looks at all of the individuals in a local community, agent based modeling allows for one agent to be created to represent each member in the community.

Aside from the number of individuals that need to be created within the model, a number of different attributes need to be assigned to each individual or agent. This is another key reason why ABM was chosen for this research project. For the

purpose of this effort, the individual agents need to be given several attributes that will accumulate values over time. These exact attributes and values will be described a bit later in the code explanation for the model. Because ABM incorporates the manipulation of these values, the model will allow for all of the agents to change in different ways, and at different rates. This allows the model to more accurately represent the real-world system, which increases the likelihood that the model will replicate the desired real-world behavior.

2. BACKGROUND

The model that was used for this research project required an extensive literature review to gather information on party affiliation. The initial literature review looked at the factors that were involved in how individuals decide which political party to affiliate with. As the research project changed, so did the need to find literature that was more closely related to the updated research question. As such, the initial literature review will not be discussed, but instead the literature relating to the updated research question will be reviewed in the following section.

The new research question focused on political party affiliation dispersion patterns, thus this was the area of literature that was reviewed. Many of the articles that were used for this research project claimed that political party dispersion patterns followed a couple of key patterns. First, the literature stated that individuals in different political parties would form tight groups within their local communities. These groups of individuals would often consist of people who shared the same political party affiliation [1]. The literature claimed that this is the case because of the influence these groups had on a new individual arriving to the neighborhood. If a new arrival to the neighborhood did not have a

high number of social network friends within that neighborhood, then the new arrival would tend to affiliate with which ever party their immediate neighbors did [5]. This affiliation tends to happen because; new individuals often tend to agree with the views of their neighbors in order to limit the degree of tension between themselves and their neighbors [2]. However, this does not tend to be the case if a new arrival to the neighborhood has several friends within their social network living in that neighborhood.

If an individual is new to a neighborhood and has several social network friends in the neighborhood, then the individual will tend to affiliate with the same political party as their friends [3]. The literature reveals that there are two possible causes for this particular affiliation pattern. First, when individuals move into a new neighborhood, in which they already have friends, the individual already has previous ties to those friends [11]. These previous ties may include the idea that a person chooses their friends based on similar political party affiliation [3]. If this is indeed the case, then individuals arriving to a new neighborhood may tend to care less about the tensions that may be created with their neighbors. This idea relies on another idea that an individual may not care as much about neighbor tensions because their neighbors are already their friends, and thus, they will affiliate with the same party [8]. The second reason individuals may tend to affiliate more with friends rather than random neighbors are; an individuals friends likely have more influence over that individual, compared to the new neighbors of that individual. If an individual receives, and accepts, more influence from their friends, then they will be more likely to follow the party affiliation of their friends [4]. The literature reveals that if this is the case, then the typical “group dispersion” pattern will not be

followed, and individuals may be outliers in otherwise tight groups of same party affiliates [4].

Overall, the literature provided a very good basis in which to start the model. Based on the literature, this research project models two separate dispersion patterns. First, the model simulates neighborhoods in which there are several tight groups of same party affiliates that tend to reside together. Also, the model simulates a situation in which there are several tight groups of same party affiliates, but those groups also include outliers, where individuals affiliate with the opposite party, as compared the majority of the group.

3. METHODS

The metacode from the research project model appears in the appendix, but a further explanation of the model may help readers to better understand the model. The ABM program Netlogo [12] was the software chosen to be used for this research project. The model initiates with 1,089 agents. The agents all initiate with a white coloring, representing the fact that these individuals do not currently have a specific party affiliation. Two of the agents are randomly chosen and given different colors, which represent different party affiliations. One agent is red, representing a Republican Party affiliation. Another agent is blue, representing Democratic Party affiliation.

Once the model begins to run, each agent looks at its surrounding neighbors and the members of their social networks, to determine the number of republicans and democrats that are present in these areas. Adjusting the appropriate sliders and choosers within the model can change the surrounding neighbors and social network variables. For every republican in an agents radius, that particular agent will add one to its “repub” variable and subtract one from its “demo” variable. This is the same for the

democratic agents, except the addition occurs on the “demo” variable and the subtraction from the “repub” variable. The system follows a similar pattern for an agent’s social network. The only change is that the social network carries more influence, so the agent will either add or subtract two instead of one. This again depends on the agents either being republican or democrat. Next, the agents look at their political threshold, this threshold is given to every agent, but the value of the threshold is random. The threshold variable is a number that determines the when a agent will changes its party affiliation. If an agent has a “repub” value that is higher than its threshold value, then the agent will become republican and change its color to red. If an agent has a “demo” value that is higher than its threshold value, then that agent will become democratic and change its color to blue. In the instances where agents have “repub” and “demo” values that are both higher than its threshold value, then the agent will affiliate with the higher of the “repub” and “demo” values and change to the necessary color. This process is repeated for every turn in the model. A turn represents one day, the model runs for a period of 365 turns, which is equivalent to a one-year period. At 365 turns the model completes and records results.

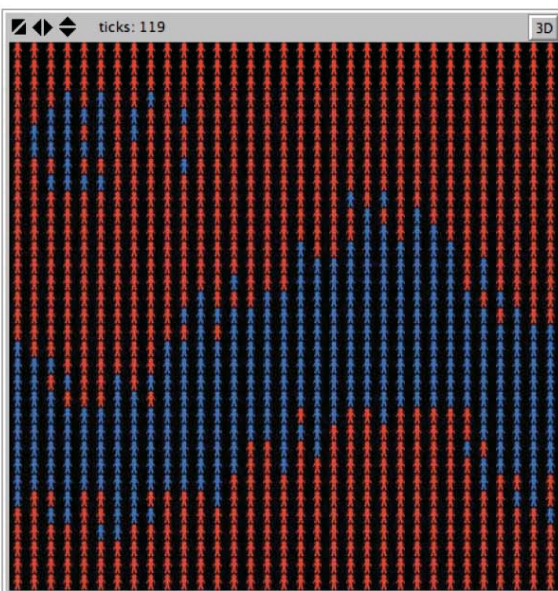
4. RESULTS

The affiliation model demonstrated many instances where all of the agents would conform to one group or another. These results were counter to the hypothesis proposed for this paper. The project hypothesis that the results would be a split dispersion, in which the project could examine the pattern in which the agents grouped, did not occur in almost all instances. The experiment had 108 different possible combinations of parameter values,

and only a couple of these parameters seem to produce the intended outcome.

The following parameter combinations were the ones that had the greatest affect on the emergent behavior of the model. First, the “starting-party-affiliates” variables was set to one. This makes the model start with only one democratic and one republican supporter. Second, the distribution of the link neighbors was set to uniform. Third, the “local-community” variables was set to three. Lastly, the “network-density” variable could be set to either one or three. These combinations of parameters were the only ones that produced results which confirmed the hypothesis.

The other 100 or so parameter combinations produced results that were not consistent with the reviewed literature. These combinations produced results in which all of the agents would conform to either the republican or democratic affiliation. These results do not allow for study of how individuals conform to party affiliation in the real world, an example would be the dispersion in a local neighborhood.



Finally, this paper will explain the process I used to determine parameter settings that produced the expected results. Since the goal of this project was to produce a model that had close to equal numbers of both republican and democrats, it was easy to determine which parameter combinations produce these results. By simply asking the behavior space, the program component of Netlogo which controls experimentation, to return the total number of republican and democratic agents, and also these percentages, comparing to determine which combinations produced these results was greatly simplified. This is also the basic way the project decided to test the relationship between the parameters and the emergence. It is known several research projects use statistical tests to confirm there results, but in the case of this project, so few parameter combinations produced the emergent behavior, that there was simply not enough information to test.

With this model, validation was somewhat complicated by the results of the experiments, but is none the less an important part of any experiment and requires explanation. The model expected an equal distribution of the number of democratic and republican supporters; the model also expected a pattern that followed a grouping effect with few outliers. The research conducted earlier in the experiment revealed that this is the most common type of distribution found in areas around the country, e.g. in a single neighborhood. This is the type of distribution the model produced, when the parameters were set to those described above.

With the results of the model collected and the prospective outcomes discussed, there are mixed feelings on the models validity. First, if the model is set to the parameters discussed in section one above, then the model produces a result that is more valid than any other. The model not

only produces a close to equal dispersion of party affiliation, but it also follows the grouping pattern that was expected. On the other hand, if the model is run using any parameters, not listed above, it produces a less valid result. Most of the time all the agents affiliate with one group or the other, and do not have an equal dispersion. Also, if there are agents that affiliate with the minority group, they are often randomly spread out and do not follow the specific pattern that expected.

Considering all of the information above, with the right parameters, the model replicates a real-world system using a simulated mechanism. Of course this is only possible if the parameters discussed in the above section are used. Otherwise, the model does not accurately represent the real-world system. Based on this, it is clear that even though the model does not produce the desired outcome, unless the proper parameters are used, it does produce a very accurate representation of the real-world system, when the right parameters are used.

A next step for this model and experimentation is refinement of the model construct and the application of the model to a much larger group of agents. The initial model was meant only as a proof of concept that party affiliation amongst a group of individuals is replicable in simulation. The results of this experiment, while not completely aligning with the literature, do support this conclusion. As such, using the same parameter settings that produced expected results, but scaling the size of the populace up, would provide greater insight into group affiliation in populaces as large as cities or countries. During this process it would also be possible to modify portions of the model to create more reliable and valid results, especially for larger populations.

5. REFERENCES

- [1]. Gross, Michael L. "Moral Reasoning and Ideological Affiliation: A Cross-National Study." *Political Psychology*, Vol. 17, No. 2 (June 1996): <http://www.jstor.org.proxy.lib.odu.edu/stable/3791813> (accessed October 4, 2009).
- [2]. Kam, Cindy D. "Who Tosses the Party Line? Cues, Values, and Individual Differences." *Political Behavior*, Vol. 27, No. 2 (June 2005): <http://www.jstor.org.proxy.lib.odu.edu/stable/4500190> (accessed October 4, 2009).
- [3]. Druckman, James N., and Kjersten R. Nelson. "Framing and Deliberation: How Citizens' Conversations Limit Elite Influence." *American Journal of Political Science*, Vol. 47, No. 4 (October 2003): <http://www.jstor.org.proxy.lib.odu.edu/stable/3186130> (accessed October 4, 2009).
- [4]. Gilens, Martin. "Political Ignorance and Collective Policy Preferences." *The American Political Science Review*, Vol. 95, No. 2 (June 2001): <http://www.jstor.org.proxy.lib.odu.edu/stable/3118127> (accessed October 4, 2009).
- [5]. Philpot, Tasha S. "A Party of a Different Color? Race, Campaign Communication, and Party Politics." *Political Behavior*, Vol. 26, No. 3 (September 2004): <http://www.jstor.org.proxy.lib.odu.edu/stable/4151352> (accessed October 4, 2009).
- [6]. Berinsky, Adam J. "The Two Faces of Public Opinion." *American Journal of Political Science*, Vol. 43, No. 4 (October 1999): <http://www.jstor.org.proxy.lib.odu.edu/stable/2991824> (accessed October 5, 2009).
- [7]. Miller, Joanna M., and Jon A. Krosnick. "News Media Impact on the Ingredients of Presidential Evaluations: Politically Knowledgeable Citizens Are Guided by a Trusted Source." *American Journal of Political Science*, Vol. 44, No. 2 (April 2000): <http://www.jstor.org.proxy.lib.odu.edu/stable/2669312> (accessed October 5, 2009).
- [8]. Best, Samuel J., and Brian S. Krueger. "Analyzing the Representativeness of Internet Political Participation." *Political Behavior*, Vol. 27, No. 2 (June 2005): <http://www.jstor.org.proxy.lib.odu.edu/stable/4500191> (accessed October 5, 2009).
- [9]. Lau, Richard R., and Gerald M. Pomper. "Effectiveness of Negative Campaigning in U.S. Senate Elections." *American Journal of Political Science*, Vol. 46, No. 1 (January 2002): <http://>

- www.istor.org.proxy.lib.odu.edu/stable/3088414 (accessed October 5, 2009).
- [10]. Vavreck, Lynn, and Constantine J. Spiliotes, and Linda L. Fowler. "The Effects of Retail Politics in the New Hampshire Primary." *American Journal of Political Science*, Vol. 46, No. 3 (July 2002): <http://www.istor.org.proxy.lib.odu.edu/stable/3088402> (accessed October 5, 2009).
- [11]. Layman, Geoffrey C., and Thomas M. Carsey. "Party Polarization and Party Structuring of Policy Attitudes: A Comparison of Three NES Panel Studies." *Political Behavior*, Vol. 24, No. 3 (September 2002): <http://www.istor.org.proxy.lib.odu.edu/stable/1558396> (accessed October 5, 2009).
- [12]. Wilensky, U., (1999), *NetLogo*. <http://ccl.northwestern.edu/netlogo/>, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

6. APPENDIX

Metacode

Initialization

Create "threshold" variable, random between 50 - 90 (determines when agents will choose an affiliation)

Create "Demo" and "Repub" variables (determines which affiliation agents will choose)

Create one turtle for every patch (ask patches to sprout 1)

Ask turtles to change shape to person

Create random republican turtle(s) (red) - (based on the slider value)

Create random democratic turtle(s) (blue) - (based on the slider value)

Ask all other turtles to change color to white

Create random social network for each turtle (based on slider value, and chooser selection)

Execution

Loop (365 ticks) - model while be based on 1-year period, each tick represents one day

Each turtle:

Perceptions

Turtles will look at their immediate surrounding neighbors (based on distribution selected), and also the other turtles in their social network.

Turtles will count the number of "blue" and "red" turtles that neighbor them and the ones in their social network.

Turtles will then add 1 to the "Demo" and "Repub" variables for each "blue" and "red" turtle neighbor.

Turtles will also add 2 to the "Demo" and "Repub" variables for each "blue" and "red" social network turtle.

Also, if a turtle is "red", then they will subtract 1 from their

"Demo" variable for each turtle in their radius and social network.

This will be the same for each "blue" turtle, subtracting 1 from their "Repub" variable for each turtle in their radius and social network.

Performance

Once a turtles "Demo" or "Repub" variable exceeds a turtles specific "threshold" variable, the turtle is asked to change its color to the one that represents their affiliation.

If a turtles "threshold" is exceeded by both the "red" and "blue" count, then the turtle will compare the "red" count to the "blue" count and choose to change to the color of the higher count.

The above procedure means that turtles can flip flop between affiliations.

Each time a turtle chooses an affiliation or changes affiliation, the turtles threshold will increase by a random amount (between 1 and 10 percent).

Once a turtle reaches a threshold of 100, the turtle will no longer be able to change affiliations, and thus must remain with the affiliation (or color) they are when the threshold reaches 100.

End loop

3.17 The National Operational Environment Model

The National Operational Environment Model A Focus on Understanding the Populace

John J. Salerno, PhD
Air Force Research Laboratory,
Rome Research Site
john.salerno@rl.af.mil

Warren Geiler
Air Force Research Laboratory,
Rome Research Site
warren.geiler@rl.af.mil

Brian Hudson
ITT Corp., Advanced
Engineering & Sciences
Brian.Hudson@itt.com

Brian Romano
Air Force Research Laboratory,
Rome Research Site
brian.romano@rl.af.mil

Jason Smith
ITT Corp., Advanced
Engineering & Sciences
jason.e.smith@itt.com

Christopher Thron, PhD
Texas A&M University –
Central Texas
thron@ct.tamug.edu

Abstract. The National Operational Environment Model (NOEM) is a strategic analysis/assessment tool that provides insight into the complex state space that depicts today's modern nation-state environment. It generates plausible futures based on the current state of an environment (baseline forecasts) and supports what-if analysis by forecasting the ramifications and consequences of potential actions within that environment. The NOEM is built upon a set of open-source, license-free capabilities, and aims to provide support for pluggable modules which comprise any given model. It currently has an extensive array of integrated modules (e.g., economics, security/rule of law, governance & social well-being – including critical infrastructure) along with a number of tools to exercise those modules. The core of the model is its populace behavior module that models the social and behavioral aspects of a populace. This paper will provide an overview of the NOEM, its functionality, and details of the behavior model.

1.0 INTRODUCTION

The National Operational Environment Model (NOEM) is a large-scale stochastic model representing the environment of a nation-state or region along with a set of capabilities which allow one to exercise the model. The NOEM enables the user to identify potential problem regions within the environment, test a wide variety of policy options on a national or regional basis, determine suitable courses of action given a specified set of initial conditions, and investigate resource allocation levels that will best improve overall country or regional stability. The different policy options or actions can be simulated, revealing potential unforeseen effects and general trends.

The goal of the NOEM is to model the daily functionality of a nation-state. A typical modern nation-state is quite complex, characterized by a myriad of interdisciplinary and interdependent components representing the various social and physical systems inherent to a specific geopolitical region of interest. These systems often involve non-stationary stochastic processes, causal loops, and non-linear relationships. The NOEM subscribes to the Diplomatic, Information, Military, and Economic actions (DIME)/Political, Military, Economic, Social,

Infrastructure and Information (PMESII) paradigm, while also leveraging nation-building and stability operations literature to develop the proper model for the conditions, circumstances, and influences commonly found in a typical nation-state.

2.0 BACKGROUND

2.1 Infrastructure

The NOEM infrastructure consists of three main components: (1) the Model Development Environment, (2) the Baseline Forecaster, and (3) the Experiment Manager. The Model Development Environment (MDE) facilitates the creation and management of the model. The MDE's Model Builder is used to create a new nation-state model or edit an existing one. Users may create new regions within a nation using the Model Builder and populate those regions with various, existing system modules. The Baseline Forecaster (BF) continually and automatically provides an updated set of projected futures for any nation-state model. Each Baseline Forecaster runs as a separate process and can service multiple NOEM Clients that have subscribed to forecasts for one or more models. Every Baseline Forecast consists of a series of replications across a single

simulation, where each replication contains a plausible outcome given an initial state vector for a nation-state model. In short, a Baseline Forecast tells us where a nation of interest is headed over time if the nation remains on its current course. The Experiment Manager (EM) enables intelligence analysts and decision makers to perform *What-If Analysis* in order to compare competing policy sets and their effect on a nation or region's stability over time, as well as reveal any potential short-to long-term ramifications of pursuing any one policy set. Users may perform "What If" Analysis to optimize a current policy or plan for future use in the event of unacceptable projected stability within a nation of interest. "What If" Analysis provides the means for users to evaluate the effectiveness of policy sets applied at various points in time with regard to a nation of interest; the idea being that the most effective and compatible policy set would be the one to consider pursuing in the real world.

2.2 The Model

The model supports the simulation and the analysis of a nation-state's operational environment. Within the MDE, the model can be configured for one or more regions, where each region is composed of a group of highly interconnected modules. These modules describe subsystems such as a region's demographics, economy, or critical infrastructure (etc.). These modules, in essence, relate to the major pillars of a nation-state based on stability operations theory (Governance, Security/Rule of Law, Economy & Social Well-Being).

Any nation-state model can be assembled by using a number of these interchangeable modules. The modules can be envisioned as entities floating within a regional space, each publishing their own outputs for use by other modules and subscribing to the outputs of other modules, all synchronized by time. That is, each module consumes and outputs values at the predefined time increment of one day. Figure 1 provides a view of the overall model.

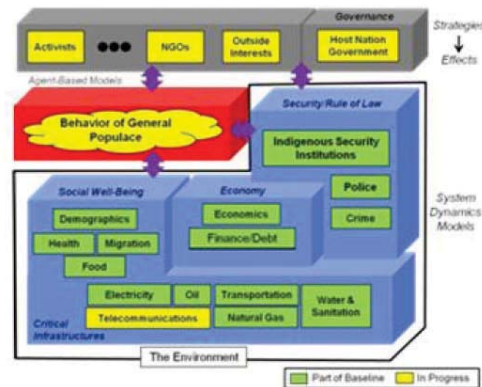


Figure 1 – NOEM Model

Security/Rule of Law is defined as protecting the lives of the populace from immediate and large-scale violence, and ensuring the state's ability to maintain territorial integrity. The Security/Rule of Law pillar is currently comprised of three modules: Indigenous Security Institutions (ISI), Crime, and Police. The ISI module is divided into: Border Patrol, Civil Defense, Facility Protection Services, and Indigenous Military.

Economy is defined as a system made up of various economic policies, macroeconomics fundamentals, free market, and international trade that exchanges wealth, goods, and resources mostly free of economic criminal activity. The Economy pillar is composed of two modules: Economics and Finance & Debt.

Governance is defined as a public management process that involves a constituting process, governmental capabilities, and participation of citizens.

Social Well-Being is defined as sustenance of life and relieving of suffering by way of humanitarian aid, best practices, human rights, essential services, and emergency response systems. The Social Well-Being pillar is composed of the majority of the modules and includes: Demographics, Health, Migration, Food, and fundamental Utilities (Electric Power, Telecommunications, Natural Gas, Oil, Transportation, and Water & Sanitation).

The Populace Behavior Module forms the heart of the NOEM model in the sense that all other modules (resources and security) are in place to support the populace. If the populace is not happy or satisfied to a certain degree, they could

easily become activists and rebel against the host-nation government. This dissatisfaction coupled with whether segments of the populace become activists depends on many factors, including their perceived hardship, legitimacy or belief in their government, their level of risk aversion, and the amount/visibility of security forces. The Populace Behavior Module, unlike all the previous modules that we have presented, is based on the Agent-Based Modeling paradigm or ABMs.

Insurgents, Coalition forces, NGOs, and Host Nation Governments within the NOEM are not modeled directly, but are characterized by the policies or strategies that they implement. Policies implemented by such groups will affect either the overall security within the environment or the services/resources provided to the people. As such, the NOEM team is looking at various Gaming Engine Techniques/Technologies that will allow one to play off these various strategies. An example area that we are currently exploring is in the area of Coalition Games.

3.0 POPULACE BEHAVIOR MODULE

In the following paragraphs, we will present the design of our initial populace behavior module, a number of experiments (based on Epstein's [1] original runs), how we modified this initial design to integrate within the NOEM infrastructure, and a number of enhancements to the model based on what we have discovered in studying the results of our initial model design.

3.1 Overall Design

We begin our discussion by describing what our agent looks like. Our agent is based on Epstein's [1] model, and similarly to his model, we compute the level of an individual's grievance based on his or her hardship and their belief in the government – legitimacy. Both values for the NOEM agents are provided by a number of NOEM modules – this is in contrast to Epstein's model, where these values are endogenous. Legitimacy is provided by the Governance module, while hardship is computed based on output generated by a collection of modules that model the

environment. Each agent is randomly assigned a set of values based on the distributions provided by the Demographics module and will have a finite life.

The number of police – or more generally, in our model, the number of personnel providing local security (including not only local police, but any indigenous military assigned to perform security detail) – are maintained through the Police module and the Indigenous Security Institutions (ISI) modules. The probability of a person (represented by an agent) being arrested is a function of the number of police and activists within a given vision (range). Based on this probability of arrest and the level of risk that an individual is willing to take on, the agent's net risk is computed. If an individual's grievance is higher than the computed net risk, they will become an activist and be added to the current number of "red" agents.

We began by reconstructing Epstein's model as described in his papers. We decided to use the George Mason University's Multi-Agent Simulator Of Networks (MASON) Toolkit [2]. The reason for this choice was that MASON is composed of a set of Java classes and is capable of being easily integrated within the existing NOEM infrastructure. In [1], Epstein provided results of eight runs using two different models. We begin our discussion with his first model. Figure 2 provides an overview of this model.

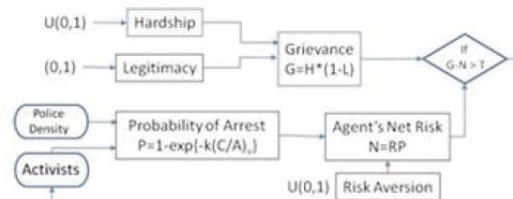


Figure 2 - Epstein's First Model

The model is divided into two groups, composed of multiple classes. The first group is the model itself. These classes include: Agent, Populace, Cop, and Epstein. Both the Populace and Cop classes are further extensions of the Agent class. The Agent class is the main part of the model. It is composed of all the rules that are adhered to by any agent. In the

first model, this class implements a populace rule, a cop rule and an optional movement rule. On each turn, each agent executes the Agent class. Each agent, based on its identity, will execute its respective rule. If it is a populace agent, it will compute its grievance level and its net risk; based on the difference between these two, the agent will either remain quiet or become an activist. If movement is turned on, the populace will then move randomly to an empty space within its neighborhood. If the agent is a cop, it will check to see if an activist is within its neighborhood. If so, the cop will move to the activist's cell and arrest the activist. The activist will be placed in jail for a random length of time and also be removed from the grid. Upon completion of its sentence, that agent will be returned back to the grid to the same neighborhood where it originated (if possible, otherwise somewhere else randomly on the grid).

The Populace class extends the Agent class by adding a number of attributes associated with each agent: Perceived Hardship, Risk Aversion, whether they are an Activist, and length of the Jail Term remaining, if arrested.

The second group of classes supports the model's graphical user interface (GUI). One of these classes, the EpsteinUI, provides the displays that allow us to visualize what is occurring on within the model. These displays include: two grids (actions and grievances) and two charts (all agents and active agents). Figure 3 provides a screen capture of the layout. The window (top right) is provided by default by MASON.

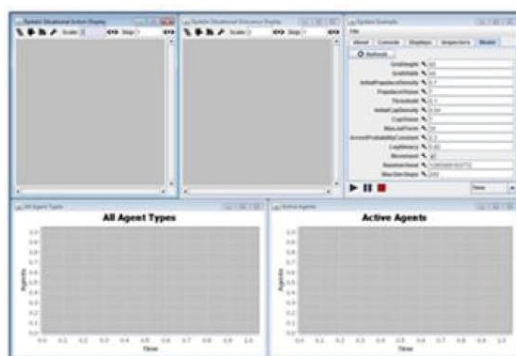


Figure 3 - Model GUI

Now that we have a working model, our next step is to verify that we are obtaining the same or similar results when compared to Epstein. For this step, we go back to his paper and attempt to reproduce each of his experiments. In his first run, Epstein examined the effect that a cop has on whether a populace, even one having a high level of grievance, will become an activist. In Figure 4, the grid on the left displays all of the populace (blue are quiet and red are active) and cops (black). The grid on the right provides us with the level of grievance for each populace. We can also see from these two grids that just because a populace has a high grievance, they may not become active as long as there is a cop within its vision and their net risk is less than a given threshold. Likewise, where there are no cops and a high grievance, we begin to see a number of the dots turning red (i.e., into activists).

Parameter	Value
Cop Vision	2
Populace Vision	2
Legitimacy	.89
Max Jail Term	15
Movement	None (Cops only upon an arrest)
Initial Cop Density	0.04
Initial Populace Density	0.7
Lattice Dimensions	40x40
Lattice Topology	TORUS
Arrest Probability Constant, k	2.3
Agent "active" threshold (G-N)	0.1
Agent Updating	Asynchronous
Agent Activation	Once per period, random order

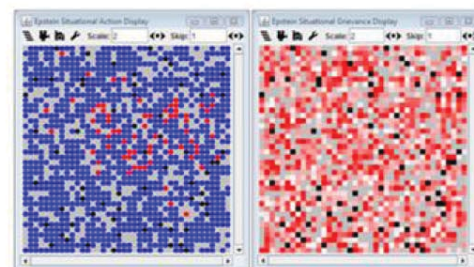


Figure 4 - Epstein's Run 1

In Run 2, Epstein investigated the rise and fall in the number of activists. Figure 5 provides the results of a run after 250 turns. As we can see from the two charts,

the number of activists will increase and decrease sharply. Since there are plenty of police to squash any rebellion (arresting the activists), the activists never achieve stronghold sustained presence. The first graph in Figure 5 provides us with the total number of activists, populace (quiet) and the number of activist that have been arrested and placed in jail. The second graph in Figure 5 not only provides us with the total number of activists but also a value that Epstein calls the “ripeness index” (blue). He believes that this indicator can identify a precursor to a possible rebellion – a possible indication of increased dissatisfaction.

Parameter	Value
Cop Vision	7
Populace Vision	7
Legitimacy	.82
Max Jail Term	30
Movement	Random site within vision
Initial Cop Density	0.04
Initial Populace Density	0.7
Lattice Dimensions	40x40
Lattice Topology	TORUS
Arrest Probability Constant, k	2.3
Agent "active" threshold (G-N)	0.1
Agent Updating	Asynchronous
Agent Activation	Once per period, random order

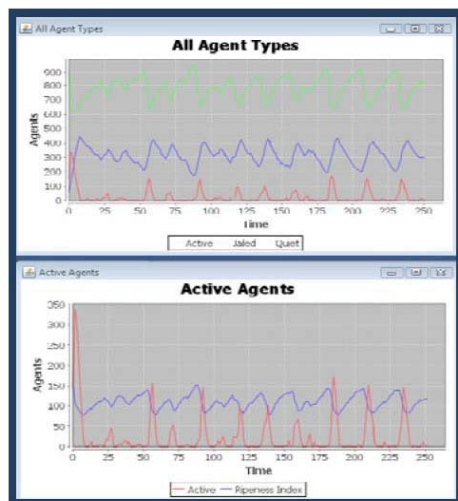


Figure 5 – Epstein’s Run 2

After we had been able to reproduce the first two runs of Epstein, we continued to

run through each of his experiments. All of our results were consistent with those reported in his papers. For brevity's sake we will next discuss how we integrated this model within the NOEM.

3.2 Integrating the Populace Behavior Module within the NOEM

In our implementation, the value for Legitimacy (of the Host Nation Government) will be provided as input by the Governance module. The number of police will be provided by the Police module, while the total number of populace will be updated via the Demographics module. The Demographics module will provide the net populace delta and will take into account those individuals that have died due to disease, murder and natural causes. It will also take into account those who have been jailed due to crimes related to violence and property. Arrests of the rebels made by the police are not accounted for by the Crime module. It will be computed by the Populace Behavior Module and fed back into the Demographics module (See Figure 6).

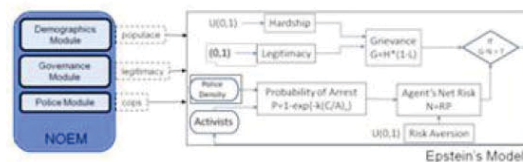


Figure 6 - Epstein's Model Integrated within the NOEM

The model was successfully integrated and initial tests were conducted to determine that similar results were obtained. To determine whether the model was still working correctly, we simply re-ran our tests a second time. All such test results were successfully replicated.

3.3 Enhancements

A particular distinction of the NOEM model is that it couples a geographical, agent-based behavioral model with a dynamical model of the social environment including security, social well-being, economics, and infrastructure. In this section, we explain some of the methods used to coordinate

this Populace Behavioral Module with the overall environmental model.

3.3.1 The Activist to Populace Ratio

How do we know how well we are doing or not doing given a specific policy? One way we propose accomplishing this is to measure the overall populace's dissatisfaction or discontent. To compute this, we take the number of populace that are currently active and divide it by the total number of those that are quiet. This is what we have titled the "Activist to Populace Ratio". Figure 7 provides an example plot.

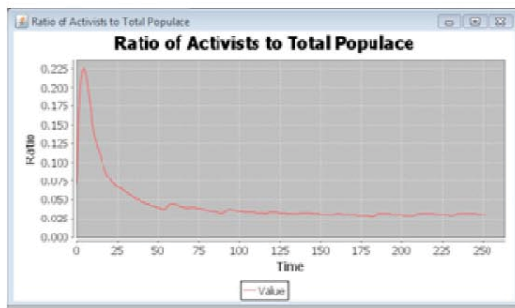


Figure 7 – Activist to Populace Ratio

3.3.2 Initialization of the Grid (Geo-registering to the Environment)

The NOEM allows for different agent types which can represent differing subpopulations, such as ethnic groups. The model accounts for geographical distribution of subpopulations by assigning agents to subpopulations according to location. Subpopulation location information may be supplied by online databases such as the Geo-referencing of Ethnic Groups (GREG) database [3]. Alternatively, the NOEM can initialize the grid if it is supplied with the centroid and approximate size of each subpopulation. Initially, the model fully populates the grid and assigns the populace agent at each location to the closest subpopulation centroid. An iterative algorithm is used to reassign points at the boundaries between regions until each region is at least large enough to contain the number of populace members in its associated subpopulation. After the initial regions are assigned, if a region contains more populace agents than required, agents are removed from that region at random until the correct number

of agents is reached. Since the NOEM also allows for migration, the subpopulation regions are not fixed throughout the course of the simulation.

3.3.3 Tying to the Environment

All agents in the grid are assigned a hardship that is calculated based on environmental conditions, and is updated as environmental conditions are updated. Hardship is calculated as a sum of terms as follows:

$$\text{Hardship} = \sum_{j=1 \dots 5} (\text{Importance of level } j) \quad (1) \\ \times (\text{Hardship associated with level } j)$$

where

- "Level" refers to 5 broad areas of need (corresponding to Maslow's hierarchy of needs) which include: physiological, safety, love/belonging, esteem, and self-actualization
- "Importance" refers to the relative importance of each of the five levels to the individual agent. The five importance values are positive weights which sum to 1.

The hardship associated with level j is in turn broken down as an additive sum of factors. For instance, the physiological hardship term ($j=1$) is computed as:

$$\text{Physiological hardship} \\ = \sum_{k=1 \dots K} (\text{Importance of factor } k) \quad (2) \\ \times (\text{factor } k\text{'s physiological hardship})$$

where K is the total number of physiological factors including air quality, potable water, food, shelter, housing, sanitation, and so on. The hardship for each individual factor is based on the model of the environment and its value to that culture to which the individual belongs. For instance, the Water & Sanitation Module indicates potable water availability by computing the average potable water supply (in liters per day per person). This availability can be translated into a hardship value by utilizing a hardship function that maps the water availability to a hardship value between 0 and 1. On theoretical grounds, this hardship function is expected to be an S-shaped curve that roughly follows the shape of a cumulative normal distribution. Other physiological hardship factors are treated similarly, and

make use of similar hardship functions. It should be noted that the hardship function is extremely flexible in that it can be closely approximate any monotonically increasing or decreasing function with up to one inflection point (that is, S-shaped or inverted S-shaped, logarithmic, exponential, or linear).

3.3.4 Legitimacy (Belief in the State)

Epstein's model requires a measure of a government's political legitimacy, which reflects citizens' inclination to support and sustain the government. There are a number of different quantitative measures of legitimacy in the literature. Typically, a legitimacy measure is the weighted sum of several indicators that reflect various societal conditions. For instance, Marshall and Cole [4] group their legitimacy indicators according to Security, Political, Economic, and Social considerations. Their grouping fits conveniently with the NOEM, which divides the environment into similar categories. The NOEM environmental variables can be used as legitimacy indicators, so that legitimacy will vary dynamically over the course of the simulation. The NOEM enables considerable flexibility in the implementation of a legitimacy measure, where various possibilities are currently being investigated.

3.4 Sample Results

In this section, we present three examples as to how the Activist to Populace Ratio can be used to compare various policy sets. In the first example, we provide the baseline (available water, 20 Litres per Capita per Day (lpcd) and tax rate, .44). In the second example, we investigate a decrease in the available water supply; while in the third example, we look at an increase in the tax rate. We would like to note here that the effect of these changes are provided as examples only and are not calibrated with a specific scenario.

3.4.1 Baseline Run

As previously mentioned, we ran the model where both the available water and tax rate do not contribute to the hardship of the populace (Figure 8).

Parameter	Value
Cop Vision	7
Populace Vision	7
Legitimacy	.82
Max Jail Term	30
Movement	Random site within vision
Initial Cop Density	0.04
Initial Populace Density	0.7
Lattice Dimensions	40x40
Lattice Topology	TORUS
Arrest Probability Constant, k	2.3
Agent "active" threshold (G-N)	0.1
Agent Updating	Asynchronous
Agent Activation	Once per period, random order
Available Water (lpcd)	20
Tax Rate	.44

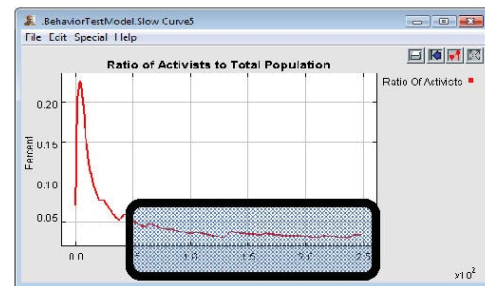


Figure 8 – Baseline Run

3.4.2 Decrease in Available Water

Our second example examines the effect that a reduction in the amount of available water (i.e., 3 lpcd) has on the populace. The decrease occurs at time 77 (days) and lasts until time 150. In this example, we can see an increase in the ratio during this timeframe and that it takes some considerable time to settle back down (Figure 9).

Parameter	Value
Cop Vision	7
Populace Vision	7
Legitimacy	.82
Max Jail Term	30
Movement	Random site within vision
Initial Cop Density	0.04
Initial Populace Density	0.7
Lattice Dimensions	40x40
Lattice Topology	TORUS
Arrest Probability Constant, k	2.3
Agent "active" threshold (G-N)	0.1

Agent Updating	Asynchronous
Agent Activation	Once per period, random order
Available Water (lpcd)	20 Time 77–150: 3 lpcd
Tax Rate	.44

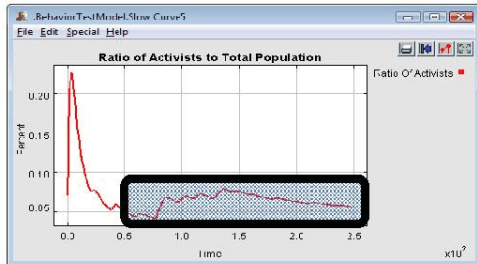


Figure 9 – Decrease in Available Water

3.4.3 Increase in Tax Rate

In our final example, we examine the effect that an increase in the tax rate has on the populace. Specifically, we increase the tax rate at time 77 (days) from 44% to 50% and keep this tax rate there until time 150. Similarly to our previous example, we do see an increase in the ratio, but not as significant (Figure 10).

Parameter	Value
Cop Vision	7
Populace Vision	7
Legitimacy	.82
Max Jail Term	30
Movement	Random site within vision
Initial Cop Density	0.04
Initial Populace Density	0.7
Lattice Dimensions	40x40
Lattice Topology	TORUS
Arrest Probability	2.3
Constant, k	
Agent "active" threshold (G-N)	0.1
Agent Updating	Asynchronous
Agent Activation	Once per period, random order
Available Water (lpcd)	20
Tax Rate	.44 Time 77–150: .50

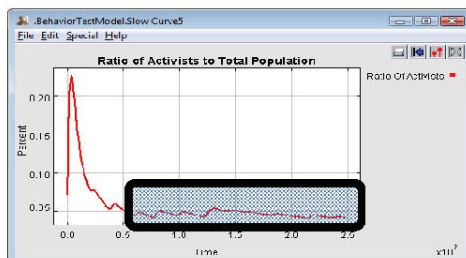


Figure 10 – Increase in the Tax Rate

4.0 CONCLUSIONS

In this paper, we have presented an overview of a tool to assist and augment decision makers and their staff with an increased awareness of the interdependencies of the components that make up a region or nation state. We provided specific details as to how we model the local populace by starting with Epstein's work and integrating it within the NOEM.

ACKNOWLEDGMENTS

This work was partially supported by Air Force Research Laboratory (AFRL), Air Force Office of Science and Research (AFOSR) under a Laboratory Intramural Research Initiative (LRIR).

REFERENCES

- [1] Epstein, Joshua M., Modeling civil violence: An agent-based computational approach, PNAS, May 14, 2002; vol. 99, suppl. 3, pp. 7243-7250.
- [2] Multi-Agent Simulator Of Networks (MASON) Toolkit, <http://cs.gmu.edu/~eclab/projects/mason/>.
- [3] Weidmann, Nils B., Jan Ketil Rød and Lars-Erik Cederman (2010). "Journal of Peace Research," July 2010; vol. 47, 4: pp. 491-499.
- [4] Monty G. Marshall, Benjamin R. Cole, "Global Report 2009: Conflict, Governance,

3.18 Exploring the Use of Computer Simulations in Unraveling Research and Development Governance Problems

Exploring the Use of Computer Simulations in Unraveling Research and Development Governance Problems

Mariusz A. Balaban
Old Dominion University
marbalarbala@gmail.com

Patrick T. Hester, Ph.D.
Old Dominion University
patrick.hester@gmail.com

Abstract. Understanding Research and Development (R&D) enterprise relationships and processes at a governance level is not a simple task, but valuable decision-making insight and evaluation capabilities can be gained from their exploration through computer simulations. This paper discusses current Modeling and Simulation (M&S) methods, addressing their applicability to R&D enterprise governance. Specifically, the authors analyze advantages and disadvantages of the four methodologies used most often by M&S practitioners: System Dynamics (SD), Discrete Event Simulation (DES), Agent Based Modeling (ABM), and formal Analytic Methods (AM) for modeling systems at the governance level. Moreover, the paper describes nesting models using a multi-method approach. Guidance is provided to those seeking to employ modeling techniques in an R&D enterprise for the purposes of understanding enterprise governance. Further, an example is modeled and explored for potential insight. The paper concludes with recommendations regarding opportunities for concentration of future work in modeling and simulating R&D governance relationships and processes.

1.0 INTRODUCTION TO RESEARCH AND DEVELOPMENT GOVERNANCE

Key tasks of governance depend on many considerations, but potential governance-level responsibilities may include strategy and operational framework design, vision or direction of organization propagation, goal setting, funding allocation, risk reduction and uncertainty assessment, effectiveness and performance assessment, decision making regarding management or lower level structures, policy adjustments, and policy adherence verification.

Governance of Research and Development (R&D) is a very broad term and has different implications for public and private sectors, however, some rules and perspectives are quite similar. The goals and objectives of these two groups are quite different, however both provide a frameworks that encourages desirable research outcomes for enhanced mission capabilities [1]. Here, it is important to understand major differences regarding how both sectors see

and evaluate their mission capabilities. Private sector governance concentrates on investments, product lifetime cycle, future planning, and sustainability [2], but the primary goal of many private enterprises is to make money. Exceptions from this rule include non-profit organizations or corporations where effectiveness aligns closer to public sector measures. Public sector governance focuses on different outcomes such as leadership, policy soundness, and innovation creation, elements traditionally viewed as specific outputs from more detailed managed endeavors. Money spent is an important factor, but the main organizational objectives do not typically focus on a return on investment paradigm.

Another context to consider is the level of R&D governance, because it may span across a society, country, industry, company, or individual endeavor within an organization. Different approaches to R&D governance can be more effective with different organization sizes and complexities. Another factor to study is the

composition of different R&D organizational structures, for example: centralized, decentralized, or mixed; this can have a significant impact on performance of the organization [3]. Assessment of an R&D endeavor is tightly coupled with organizational goals. Input in the form of goal setting will direct the R&D process through execution of projects, and ultimately it will feedback to the governing cell in the form of, for example, higher profits, increased market share in the private sector, lowered crime rates, decreased cancer-related deaths, and an increase of the approval rating of a given organization in the public sector. A formulated governance strategy establishes a context for R&D entities and a consistent basis for funding requests, adjustment of programs or possible outsourcing to accomplish organizational goals. It is also an enabler of generating new ideas for future projects, which would help to fill the gap between current state, and desired organizational objectives.

The remainder of this paper is organized as follows. Section 2 discusses the applicability of M&S techniques to R&D governance. Section 3 briefly presents the most common M&S paradigms. Section 4 discusses the problem of choosing a modeling paradigm in the context of R&D governance. Section 5 presents a sample model. Finally, Section 6 provides some conclusions regarding this research.

2.0 NEEDS FOR MODELING IN R&D GOVERNANCE

Decision-making in complex systems without sufficient information and understanding could lead to catastrophic results. Further understanding of complex systems may be achieved through M&S techniques. Frequently, available data do not translate into valuable information that facilitates understanding. When investing a significant amount in a research project, one should first know that a proposed solution

has a high chance of success, and if the proposed solution is actually addressing the concern [4]. Another problem can be discovered while looking at reports of government agencies where many voluminous reports do not provide clear information on the bigger picture context.

Measures of "target met"-type results for particular research endeavors do not provide a sufficient way to validate a decision-making process and its evaluation. On the other hand, this situation is likely due to a lack of knowledge on how to successfully describe R&D processes, and their relations at a sufficient level of abstraction. The problem can be instantiated into the scope of conceptual understanding to see how this system works. If one knew how the system behaves, a model could be built out of this knowledge to aid in decision-making process. This sounds straightforward, but in fact, it is not. There are many parts of the system influencing each other and it is difficult to describe these relations in a mathematical formalism. Further, human-in-the-loop concerns are still a big challenge to model efficiently, and when dealing with innovation it is obvious that not only qualifications of scientific staff, but also other factors such as their motivation, character, and needs, which could influence the culture and performance of the working group are important. Therefore, evaluation of the process of innovation generation can be subjective and a decision-maker concerned with organizational governance should be aware of that. M&S practitioners are working diligently trying to fill this gap, because common analytical techniques cannot.

Agent Based Modeling is one of the most promising techniques that could generate an understanding of innovation phenomenon and emerging group behaviors. A question often asked is how to assess and measure R&D effectiveness. One way of approaching the problem is to focus on internal and external factors that

facilitate effectiveness and not on the innovation product itself, especially when assessing a long-term project, where it is difficult to gauge the progress of research. Another way is to find variables directly coupled with outcomes from an R&D endeavor, keeping in mind that some variables can be measured, but some can only be observed as relative trends. Length and difficulty of the R&D endeavor can have a significant impact on accuracy of the prediction, therefore overall assessment as well. When considering the length it is apparent that long-term R&D endeavors should be more difficult to appraise and the probability that one will conduct an erroneous assessment is higher. By employing modeling techniques, one could address some of the problems mentioned here. Modeling the innovation process and its supporting activities can be achieved using modeling techniques. They can alleviate a problem of R&D effectiveness assessment and help to evaluate trends and conduct "what if" analysis on different styles of decision-making and different policies of R&D governance that oversee innovation generation. Successful conclusion of R&D innovation generation is not the end of the lifecycle process. Another responsibility of R&D governance is to incorporate a new product or concept into a market or its functional target place. This task can be very complex and traditional methods may not suffice. M&S can mimic time-related variables and environmental relationships related to new inventions to optimize the process of integration to maximize outcomes, and to avoid pitfalls that could possibly occur.

Another area where M&S can be invaluable is assessment of future R&D direction. Conducting analysis of R&D effectiveness using modeling techniques in the context of organizational mission can reveal serious strategy flaws and could help to avoid them. M&S methods are relatively new approaches, as a choice for aid in decision-making and evaluation of R&D governance. Authors were looking into areas where M&S

is already more advanced to analyze options in applying robust knowledge to R&D governance.

The most prevalent area of M&S is associated with, and sponsored by, military organizations. In [5], authors used ABM to model Political, Military, Economic, Social, Information and Infrastructure (PMESII) entities to mitigate the impact of disasters and military operations. This model has been used to represent the rebuilding and expansion of critical infrastructure in Afghanistan. The ability to gain insight into the process of reconstruction and development can be a very valuable tool for a decision-maker in enterprise or government structures. Because of the multi-dimensional scheme of the model, the authors believe that it would be possible to apply this approach to R&D governance.

Adding modules that would allow agents to model innovation generation and knowledge sharing is possible with modeling by connecting this lower level to middle and higher-level agents, e.g. department, organization, and society. A model could allow the analysis of the robustness of policy choices under different scenarios. This approach can be used in the public as well as private sector. Adding more complexity to the model would theoretically allow for modeling human evolution as a function of R&D progress. Because the progress of R&D could be an enabler for actual change of the system and can spur achievements of higher objectives of system of systems, it can be invaluable to be able to model this process. As a futuristic perspective, such a model can provide valuable information on R&D as a stabilizer or destabilizer of the world, and perhaps allow answering the following questions:

Is advancement of R&D proportional with civilization advancement?

Is our direction of R&D advancement taking us where we want to be?

Because of its complexity, currently only large organizations can handle expenses

related to investing in complicated models conveying many levels of abstraction, but this can change as the field of M&S advances, providing for easier model reuse and better interoperability. The following section discusses the most common modeling paradigms to familiarize the reader with what approaches are available.

3.0 MODELING PARADIGMS

In this section, the authors would like to provide quick overview of three popular and often-used-by-practitioner M&S methodologies. Understanding principles of them is an important step in understanding their possible usage, as discussed in the next section.

3.1 System Dynamics

System dynamics (SD) models consist of feedback loops in the form of differential equations that provide for building relations between variables. SD is useful for studying complex nonlinear systems, especially finding cause and effect relationships [6]. From the practical perspective on how to build model, one should know about stocks, flows, auxiliary variables, and constants as the main blocks providing for metaphors of the complex system. The initial phase of building SD model is creating a causal diagram. A causal diagram represents the most important elements of a system and relations within the system in form of links that end with arrows indicating which element influences which. In addition, you can denote a "+" or "-" on the flow to specify a positive or negative relationship. The positive one means that second variable follows the direction of change of the first variable, and vice versa. Links with variables form loops, and they can be positive or negative as well. The negative loop can exist only if the number of negative relationships is odd. Having at least one negative loop is a minimum requirement to stabilize the system. Models built with this methodology can help in framing problems, revealing dynamics related to change

imposed on the system. Models are typically used to show trends of relationships and not precisely computing specific values. This methodology can be an invaluable tool in assessing a big picture problem, testing alternative policies and strategies at the governance or enterprise level.

3.2 Agent-Based Modeling

The world is full of interactions between different types of systems and many of them are not well understood. Agent-based modeling (ABM) is a methodology that aims to capture many types of interactions by using computer created entities called "agents". These agents are assigned attributes, rules of behaviors, mimicking and carrying sets of interactions to gain an understanding of the real system, and providing for emerging behaviors. ABM can be used to model perception, autonomous and group behaviors, goal setting, a static or dynamic environment, the ability to mimic inference and adaptation, and social interactions among other possibilities [7]. ABM in R&D governance allows for the examination of different types of relations between individuals, groups, and organizations, with the notion of independence of agents, when modeling for example open innovation or supplier-customer relations. These concepts clearly could take advantage of ABM to reveal some of their intricacies. For these reasons, it is popular to use ABM to generate hypotheses about system behavior. From the scientific point of view, this is a very powerful feature, but there are some practical limitations related to an inability to logically validate such models. Consequently, verification and validation (V&V) is an issue that needs to be resolved to take full advantage of ABM. One way of partially alleviating this drawback is by proving a hypothesis false.

Complexity and dynamic behavior of systems encourages the use of ABM for investigation of adaptive systems, which are often non-linear and chaotic. Furthermore,

provided mechanisms lead to model evolution of adapting agents [8]. This capability sets the ABM methodology apart from other methodologies discussed in this paper. It is worth mentioning that relatively high complexity can be achieved by imposing relatively simple interaction rules on agents. There are many platforms for implementing AMB, e.g. NetLogo, Swarm, Mason, Repast, and AnyLogic. The choice of one of these platforms would greatly depend on the complexity of system that one would like to model and the problem that the constructed model is supposed to address.

3.3 Discrete Event Simulation

In a discrete model, changes in a virtual system can occur only at separate points in time. These changes are called events, and everything in the model is related to them in one way or another. For example, an event could be an arrival of a R&D project for its approval, then decision made would mark next event, following with outcome event from the project e.g. success or failure. Discrete Event Simulation (DES) model consists of entities, which in an R&D environment could be represented by R&D projects flowing through model logic. Events are stored in a calendar, which contain information that allows the model to be executed in accordance with its logic. The central idea of DES is that variables of the model will not change between successive events.

In addition, understanding the concept of a queue is often necessary to the understanding of DES. A queue can be visualized like a line in a store, and can be modeled to mimic the system's real queue, which may have limitations for the number of elements that can fit into it, and can have different rules reflecting the priority of leaving it by the stored elements, e.g. a first in first out relationship.

Essential components of DES are represented by resources, which relate to

personnel, equipment, etc. of the real system. They will be used by entities while going through the process. DES is often used to capture stochastic behavior of a system [9, 10].

4.0 MODEL CHOICE

One can model R&D governance at various levels of abstraction, complexity, and fidelity depending on purpose of a model [11]. It is important to understand some basic limitations of M&S to proceed with this discussion further. One needs to remember that modeling activity is not free, but can, if executed properly, save a great deal of money and time.

Model development and validation often requires a significant amount of time and resources. No one model answers all questions, so it is important to evaluate cost effectiveness when deciding to use M&S techniques, meaning, if one can address a problem with simple deterministic and analytical calculations, then M&S is most likely not needed. Analytical Methods (AM) are usually quicker and cheaper. AM are often used to get an exact answer to a problem, usually where there is a direct and known relationship between input and output.

Short length term R&D endeavors should be especially good candidates for use of AM in effectiveness assessment at the managerial level. This could be done, for example, by combining R&D into business process. In some cases, AM allow for quick estimates for a problem, which then needs to be addressed more thoroughly with a M&S paradigm, as AM would not allow to capture the bigger scope or an appropriately complex problem.

Generally, the choice of a modeling technique depends on what features of a system can be effectively manipulated with different methods. Aggregation level is another factor that will influence the choice of modeling technique. It is important to

understand that a lower level of aggregation does not necessarily translate into lower risk associated with, e.g., prediction. At the lower level of abstraction of R&D governance one can focus on innovation generation, knowledge management, process of innovation control and supervision, and project management scope. These could provide an instrument used for medium and higher level structures, or can be studied on their own as separate models. Modeling emerging behavior using ABM could potentially be a method to use at the lowest level of abstraction of R&D governance. One could get insight into new order, new patterns, and structures of R&D by learning from the interaction of individual agents or groups.

One possible approach to implementation would consist of distributed agents in the form of humans, information storage, or groups working together, with attributes such as goals, knowledge, and imposing adaptive tension on one another [12]. Outcomes from the emerging behavior can be greater than the sum of its components, e.g. knowledge generation; tacit knowledge depends on interaction and direct experience, whereas explicit knowledge could easily be transferred across distance. Modeling a spectrum of interactions would allow for a better understanding of complex emerging behaviors, along with quantification of expected knowledge increase. Different structures at different levels in enterprise or government along with imposed interactions could, for example, increase chances for innovation generation. Emerging behavior capabilities of ABM can provide for building new theories, which in turn can be translated into better structures, encouraging motivation, and knowledge sharing in an organization that enables creativity. As discussed, it is feasible that ABM can be mostly suitable for theoretical development, which should be later validated using methods that are more empirical, as ABM is generally difficult to validate. ABM can also be suitable at higher levels of R&D governance modeling, where

individual agents mimic projects, organizations, societies, regions, and countries, and agents' independence can provide for different perspectives comparing to one obtained from, e.g., the SD method.

DES seems appropriate to use when looking at R&D governance from the processing perspective [13]. Every big enterprise or government entity has to manage many R&D projects, evaluate them, and properly allocate resources. DES allows for modeling of R&D endeavors throughout their stages in a life cycle. In DES, R&D projects can be represented as entities going through their life cycle, starting from project approval request, which can end up as a rejection, then the development phase through stages at which it is possible for a project to be rejected from continuation due to, e.g. a lack of funding. This would allow the transferring of resources allocated to this project to another project, increasing its chances of success. Finally, a project's output needs to be incorporated into production, then into market, or another environment than for which it was initially conceptualized. Additionally, one cannot forget about a product's maintenance tasks throughout its life cycle. The R&D process calls for a method that can easily incorporate its stochastic character, and DES provides that capability. The ability to mimic the logic of the organization, and evaluate alternative structural alignments, for example, centralized versus decentralized, makes DES especially valuable. In addition, the stochastic character of R&D endeavors makes DES a very powerful tool that could be used for assessment of, e.g., outsourcing versus "in-house" R&D, and prediction of estimated return of investment. DES can exhibit a detailed level of abstraction depending on the modeler's need, and it is flexible and easy when managing different levels of aggregation. Moreover, many commercial DES packages provide techniques that can mimic different methods like SD and some even ABM.

SD is a method that can be very useful, especially at the enterprise modeling level when trying to capture big-picture or global scope of the problem. SD should be used where there is an existing feedback loop between model variables, and it is hard to explain the relationship by intuition itself, which often appears in the representation of R&D governance. Problems of choosing between various strategic options can be addressed with SD, as it is often hard to model high-level of socio-political phenomenon providing exact numerical information. This method could be particularly appropriate to a long-term model. SD structure is based on differential equations, but most of the commercial simulation packages allow for the incorporation of stochastic behavior into a model. Using this method for testing policies and future trends can be invaluable when trying to plan in long term, which is specifically related to R&D governance type decision-making. Modeling competing environments allows for a comprehensive perspective of what is needed to obtain desired outcome. An important advantage of SD is that data used to drive a model do not have to be ample or even direct. Often one would like to learn the system and restore its behavior without direct knowledge of it, by building relations that are only correlated with it or related to this phenomenon. Another area where SD may be appropriate is to learn about how significant the roles of parts of a complex system are in the functioning of the whole body, e.g. not only how outsourcing affects a single project outcome, but also what are the long time implications.

To manage different levels of abstraction one could potentially increase effectiveness by using nested, also known as, hybrid, models [14]. These models use different methodologies that can be arranged hierarchically depending on type of problem the model is to address. Multiple layers of ABM, DES, and SD methods could be constructed, allowing communication between layers and so influencing each

other. Some structures can be also embedded by another if that reflects actual system structure. M&S practitioners often combine ABM and SD. These two methods can work well together and provide good combination of, e.g. lower level ABM with SD capturing a larger view of the system. Alignment of ABM-SD can be multi-layered, for instance, some part of agent behavior (internals of agent) can be described using differential equations from SD, then the environment level can be described in ABM terms with agents at different level of aggregation, and on top of that SD world-view, capturing global dependencies. Some of commercial as well as free software packages allow for that kind of configuration e.g. ANYLOGIC, and NETLOGO. In Figure 1, the authors propose an example of how a configuration of a hybrid R&D governance model could be assembled.

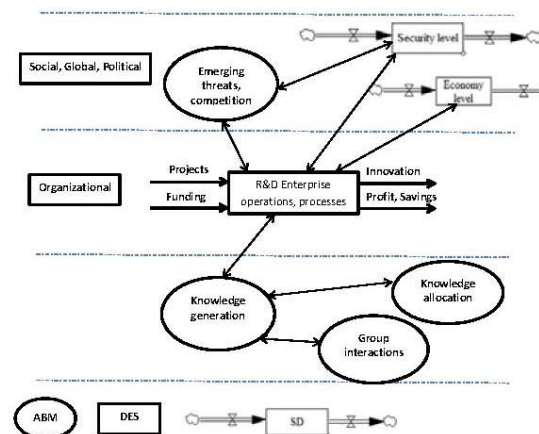


Figure 1. One possible configuration of a hybrid model of R&D governance

5.0 MODEL EXAMPLE

The estimated number of illegal immigrants ranged from 8.5 to 11.8 million people between 2000 to 2009 [15]. While this can be an under- or over-estimation, many unofficial sources estimate it as high as over 20 million people, which if true could cause a significant problem in calculations of many

statistics and biasing the true picture of the U.S. border security problem.

Figure 2 presents a simplified conceptual example which captures relationships between two competing environments, (represented as capabilities: Department of Homeland Security (DHS) capabilities and criminal force capabilities), in relation to U.S. border security.

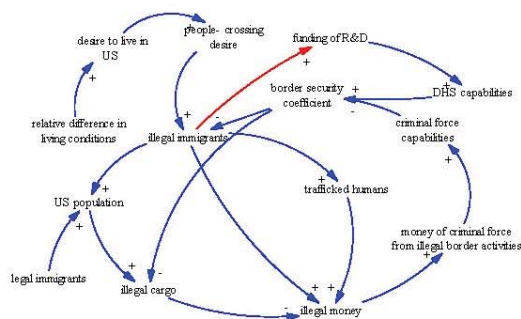


Figure 2. Conceptual model of border security

Money is used as a main metric in this model, used for both funding an increase in DHS capabilities and to provide illegal money to a criminal force. A defined coefficient of security influences number of illegal immigrants entering the U.S. and the amount of illegal cargo. The model allows for estimating the number of unauthorized immigrants in the U.S.

The model has the following data-driven variables: population growth rates, legal immigration rate, Gross Domestic Products (GPDs) of the U.S. and the world rates, and federal R&D funding for the U.S. Figure 3 presents a detailed view of the model. The model correlates the U.S. GPD with total spent money on public R&D and further on DHS. This can allow analysis to be conducted on both the country, as well as departmental level. The U.S. GPD further influences the relative ratio of difference in living conditions between the U.S. and the world. The R&D portion of the model is built on the concept of knowledge generation by both environments: one being the R&D

environment, where conducted or contracted-out research by DHS is supposed to help eliminate illegal activity around the U.S. border, and the second environment represents knowledge of the criminal force to continue with their illegal endeavor.

There is an assumption on the time delay between the reaction time needed to discover new techniques by both competitors. The authors hope that there is an effective way of breaking this pattern, but for a simplification, breaking the pattern is not considered in this model. Future study of this phenomenon using agent-based modeling approach is foreseen, and it would possibly allow for assessment of how fast a criminal force can adjust to changes imposed by border security advancements and vice versa. An increased border security coefficient can decrease effective smuggling and illegal border crossing, which in turn decreases money flow used by criminal force. The last part of the model focuses on different groups of people. Here, the key elements for our analysis are represented by the number of illegal immigrants and the level of human trafficking. The basis for obtaining those outputs is the assumption that the majority of illegal immigrants are those who cross the borders for economic reasons. Hence, the authors were looking at estimating differences between (GPDs) as a factor of estimating the number of people with a desire to live in the U.S. Additionally, human trafficking is strongly correlated with illegal immigration [16]. This model was partially validated by comparing estimated data from [15] and model output as shown in Fig. 4. The model shows potential for gaining valuable insight into border security, including estimation of the funding level necessary to lower illegal activity related to the U.S. borders.

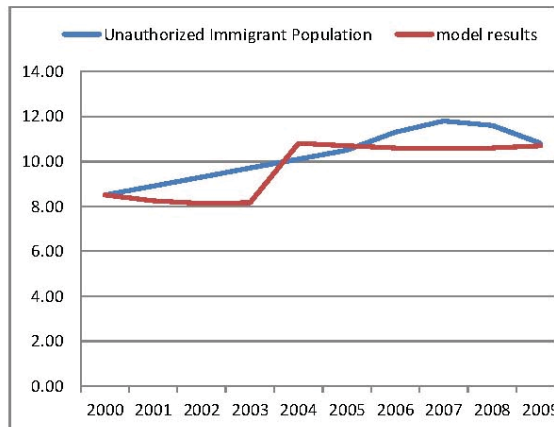


Figure 4. Partial validation of the model

6.0 Conclusions

This paper presents insights into governance of R&D. The authors moved from theoretical knowledge of R&D governance, through description of popular approaches available for M&S practitioners, ending up with discussion on how one could model R&D governance. In addition, authors showed an example of SD model that could be used by DHS or other similar agency to determine R&D funding. The constructed model allows for estimation of the projected number of unauthorized immigrants given a current or augmented rate of R&D funding. Future work is foreseen that would incorporate hybrid modeling to better capture the complexities of R&D governance.

7.0 REFERENCES

- [1] Keating, C., et al. *R&D system governance: an emerging perspective*. in *R&D Management Conference: Information Imagination & Intelligence in R&D Management*. 2010. Manchester, UK: University of Manchester.
- [2] Pearce, A.R., K.L. Sanford Bernhardt, and M.J. Garvin. *Sustainability and socio-environmental systems: modeling total cost of ownership in capital facilities*. in *Winter Simulation Conference*. 2010.
- [3] Argyres, N.S. and B.S. Silverman, *R&D, organization structure, and the development of corporate technological knowledge*. *Strategic Management Journal*, 2004. 25(8 9): p. 929-958.
- [4] EIRMA, *Assessing R&D effectiveness*. 2003, EIRMA.
- [5] Chaturvedi, R., et al. *Continuous Validation Framework: A Case Study of SEAS and Afghanistan*. in *The Interservice/Industry Training, Simulation & Education Conference (IITSEC)*. 2008: NTSA.
- [6] Garcia, J.M., *Theory and practical exercises of system dynamics*. 2006: MIT Sloan School of Management.
- [7] Bonabeau, E., *Agent-based modeling: Methods and techniques for simulating human systems*. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. 99(Suppl 3): p. 7280-7287.
- [8] Middleton, V. *Simulating small unit military operations with agent-based models of complex adaptive systems*. in *Winter Simulation Conference*. 2010.
- [9] Kelton, R.P., R.P. Sadowski, and D.T. Sturrock, *Simulation with Arena*, ed. 4th. 2007: New York: McGraw Hill.
- [10] Law, A.M., *Simulation modeling and analysis*. 4th ed. 2007: New York: McGraw-Hill.
- [11] Hester, P.T. and A. Tolk. *Applying methods of the M&S spectrum for complex systems engineering*. in *Emerging Applications of M&S in Industry and Academia (EAIA)*. 2010. San Diego, CA: ACM.
- [12] Kitaygorodskaya, N. *A framework for agent-based modeling of communication and knowledge networks in corporate R&D*. in *Computational Cybernetics (3rd International Conference)*. 2005: IEEE.
- [13] Bodner, D.A., W.B. Rouse, and M.J. Pennock. *Using simulation to analyze R&D value creation*. 2005: Winter Simulation Conference.
- [14] Chahal, K. and T. Eldabi. *Applicability of hybrid simulation to different modes of governance in UK healthcare*. 2008: IEEE.
- [15] Hoefer, M., N. Rytina, and B.C. Baker, *Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2009*. 2009, DHS Office of Immigration Statistics. p. 1-8.
- [16] Clawson, H.J., N. Dutch, and A.a.L.G.G. Solomon. *Human Trafficking Into and Within the United States* [web site] 2009 [cited 2011 July, 10, 2011]; Available from: <http://aspe.hhs.gov/hsp/07/humantrafficking/LitRev/index.shtml#Trafficking>.

3.19 An Agent-Based Model Of New Venture Creation: Conceptual Design for Simulating Entrepreneurship

An Agent-Based Model Of New Venture Creation: Conceptual Design for Simulating Entrepreneurship

Mike Provance
Old Dominion University
mprovanc@odu.edu

Andrew Collins
Old Dominion University
acollins@odu.edu

Elias Carayannis
The George Washington University
caraye@gwu.edu

Abstract. There is a growing debate over the means by which regions can foster the growth of entrepreneurial activity in order to stimulate recovery and growth of their economies. On one side, agglomeration theory suggests the regions grow because of strong clusters that foster knowledge spillover locally; on the other side, the entrepreneurial action camp argues that innovative business models are generated by entrepreneurs with unique market perspectives who draw on knowledge from more distant domains. We will show you the design for a novel agent-based model of new venture creation that will demonstrate the relationship between agglomeration and action. The primary focus of this model is information exchange as the medium for these agent interactions. Our modeling and simulation study proposes to reveal interesting relationships in these perspectives, offer a foundation on which these disparate theories from economics and sociology can find common ground, and expand the use of agent-based modeling into entrepreneurship research.

1.0 INTRODUCTION

The field of entrepreneurship has developed a productive momentum in studies that empirically demonstrate growth factors and rates, relationships between entrepreneurs, industry participants, and institutional agents, and macro-level outcomes, such as regional economic vitality, associated with entrepreneurial success. With this momentum, complexity has followed. Rather than a convergence around a core set of theories explaining entrepreneurship, the field has developed into a diverse and loosely-connected confederacy of theoretical lenses [1]. This diversity has led some scholars to call for a systems perspective of entrepreneurship because so

many different participants are involved in this economic process.

2.0 A HETEROGENEOUS SYSTEMS PERSPECTIVE OF ENTREPRENEURSHIP

Two theoretical foundations — agglomeration, and entrepreneurial action — have increasingly been applied to these systems-based explanations of entrepreneurship. Agglomeration emphasizes the environmental factors within which entrepreneurship happens, and the great inertia involved in acquiring resources from the environment that leads entrepreneurs to set up their businesses as close as possible to their prior employers and other local resources providers.

Entrepreneurial action places that evolution of the entrepreneurial ecosystem in the hands of the entrepreneurs, who seek out resources to maximize the potential for their organizations. The challenge with using either of these theoretical anchors is that both conditions exist in any entrepreneurial system. This poses both phenomenological and methodological problems with developing a better understanding of the formation processes of new ventures within economic environments.

Recent work has sought to discern the differences in effect from both theories and provide a methodology for better understanding the systems perspective of new venture formation [2], [3], [4]. This study continues the development of a simulation model for new venture formation that will clarify the relationship between agglomeration and entrepreneurial action. An overview of the competing theories is presented here along with a model for the design of an agent-based simulation to test the relationships between these theories with respect to new venture formation.

The systems perspective of entrepreneurship builds on a set of assumptions and observations that, in any economy, there are domains of activity that spur the emergence and growth of new firms to act on market opportunities. The domains comprise participants and interactions involving socio-economic, institutional, political, and technological factors — the results of which evolve into a heterogeneous complex system. Founders instigate the process by attempting to exploit novel ideas that they believe address market opportunities. These efforts may be hampered or helped by policymakers, regulators, economic development offices, and other political participants. Universities may be involved in the transfer of knowledge that leads to opportunities for exploitation. Financiers and suppliers enable that venture to gain traction, and alliances may be formed with others in the market to expedite product or service

development. Finally, customers emerge who are willing to try something new, completing the cycle of innovation.

Explanations for these dynamics have been drawn along two lines: agglomeration of knowledge exchange to produce concentrated economic activity, and entrepreneurial action that results from founders' asymmetric access to economic and technical knowledge of market opportunities. The main premise of agglomeration is that jurisdictions — primarily cities and regions — absorb extraordinary pecuniary gains through the localized spillover of knowledge from incumbent firms. Further, this spillover acts as the source of new venture creation since it leads potential entrepreneurs outside the boundaries of these firms to acquire the knowledge inexpensively and develop novel product and service offerings from it.

An important source of connection between both theoretical positions is the role of institutional forces in shaping new ventures during the early stages of their growth [4], [5]. In Reference 4 (p.2), the role of institutional action is characterized as perhaps even more important to new venture creation and survival than the competitive dynamics more often associated with entrepreneurship:

The actions most typically associated with entrepreneurs are competitive ones in which new technological pathways are constructed or new models for rent appropriation are devised. However, actions against institutions represent complements to the competitive ones that may shape a new venture's growth. In fact, issues of institutionalization and legitimacy may dominate other forces. As Boulding [6] suggests: "It can be argued indeed ... that the dynamics of legitimacy ultimately dominate all other elements of social systems." Certainly, most scholars agree

institutional forces are fundamental elements of entrepreneurship [7]. What remains less well understood is how new ventures combine competitive and institutional actions to improve the likelihood of success.

Institutions and the dynamics between these institutions and other participants are the foundations of entrepreneurial activity in economic markets [8], [9]. New ventures may be catalyzed through policy choices regarding intellectual property, financial liquidity, and competitive dynamics. Other policies may act against them, retarding their growth or providing obstacles leading to delays in creation or failure. Thus, a simulation model designed to evaluate new venture formation from distinct theoretical perspectives must account for both competitive *and* institutional forces.

3.0 DESIGN OF AGENT-BASED MODEL OF NEW VENTURE FORMATION

It is important to capture only the important aspects and variables of a system when constructing any agent-based model. This viewpoint is in-line with the Law of Parsimony (or Occam's Razor), which is the well-established principle used with social sciences that states "All things being equal, the simplest solution is the best or Entities should not be multiplied beyond necessity" [10], [11]. This approach does produce problem when building a model around two competing paradigms, namely: business growth happens due to incumbent firms, or business growth happens due to entrepreneurs bring new knowledge to an area. The agent-based model outline in this section captures the essential elements of both paradigms under consideration for the research question.

Following the Law of Parsimony, the key features of the problem-space under consideration should be captured within this

model. A list of these features is given here:

- Businesses form the agents within the model; these include incumbent firms and entrepreneurs.
- Businesses are able to form alliances with one another; these range from informal communication to company mergers.
- There is a geographical space where businesses are physically based; these spatial regions are affected by local laws, resources and taxes.
- Business sell products based on their current knowledge; these products are could be physical or service-based.
- Customer demand determines the survival of a product and ultimately the business with which it relates. A highly demanded product allows a business to grow.

This list does not represent all aspects of business modeling with key features like capital investment and production being missing. This list does, however, allow for the basis of a model that will investigate the research question.

3.1 Model Setup

The model was built around the existing entrepreneur literature and its examples. Using the existing theory to construct a model is called an axiomatic approach as opposed to a realism approach that would model based around real-world data [12]. The advantage of the axiomatic approach is that information required is readily available, i.e. through journals articles. The disadvantage this approach is that any theories acceptance, by the scientific

community, might change over time making the model invalid as well. Given the difficulties of collecting data on entrepreneurs, the realism approach is simply impractical for the research. Using the axiomatic approach, the ENTRABMS agent-base model was developed. The ENTRABMS satisfies all the elements given in the list above. There are some other important features that ENTRABMS has, namely: adaption and multiple spaces.

The authors in Reference 13 argue that see any true emergent behavior within an ABM, the agents must be able to adapt [13]. Emergent behavior is purpose of using ABM and has been argued that it is the foundations of a new scientific method [14], [15]. Emergent behaviors are outputs from a simulation that were not immediately obvious from the on-set but give insight to the analysis about modeled system. As this is kind of result that is required from this research, including adaption within the ENTRABMS is essential. The agents within ENTRABMS are able to adapt through their ability to move through the knowledge space.

The knowledge space within the model represents the abstract landscape with which both the agent-businesses' products and customer demand reside. This space is analogous to a rocky coastline where the water levels represents the customer demand and rock surface represents the currently unknown knowledge. A diagram in Figure 1 shows this analogy. The business-agents could be considered to be crabs that can move around the rock-surface (product change) or burrow into the rock surface (engage in product research).

To understand this knowledge-space, consider the example of smart-phones. There clearly was a customer demand for smart-phones but during the late nineties the computing and cell-phone technology did not exist to produce the smart-phone product. Within our analogy, "smart-

phones" could be considered to be the rock that sits between "computing" rock and "cell-phone" rock. As business-crabs burrowed into "computing" rock and "cell-phone" rock, this opens up the "smart-phones" rock and customer demand flows into it. Now eventually a business-crab will move into that space and take advantage of this demand by producing smart-phones. The knowledge space is not the only space with the ENTRABMS model. The agents in the model exist in the "real-world" geographical space as well. This need for two spaces is due to the research question requirement to investigate incumbent firms in an area and entrepreneurs bring new knowledge to an area.

This is not the first ABM that has agents existing in multiple spaces. An insurgency model presented at last year MODSIM World conference contain insurgent-agents that existing in both a social network space and an "inclination" space [16]. The model was developed using the U.S. Marine Corps War-fighting Laboratory's Pythagoras ABM package [17].

3.2 Implementation

The business-agents exist in two different spaces, they have ability to move around and they can make connections with other business. With this knowledge in mind, the ENTRABMS was built. The model was constructed using Repast Symphony. Repast Symphony is a Java-based agent-based modeling environment [18]. It was chosen because it can handle agents existing in multiple spaces and gives the flexibility of programming which is required for such a complex model as ENTRABMS. The difficult issue that now must be resolved is what initial configurations should be used to adequately investigate the research questions.

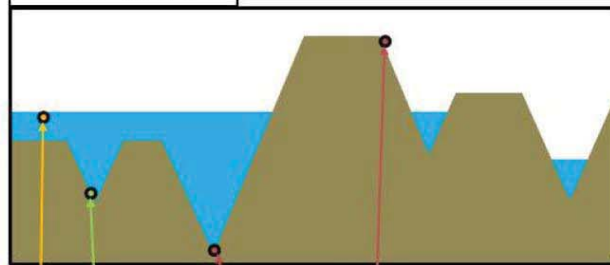
ENTRABMS MODEL

Businesses are able to move within the knowledge space through partnerships and innovation, which allows them to dig deeper.

Businesses have present in both the knowledge and geographic space

Knowledge Space

Knowledge space represent the abstract location of the business' products.



The knowledge space is analogies to rocky coastline. The water represent the customer demand and the sand is the undiscovered knowledge.

Geographic Space

Geographic space represent the physical location of the businesses

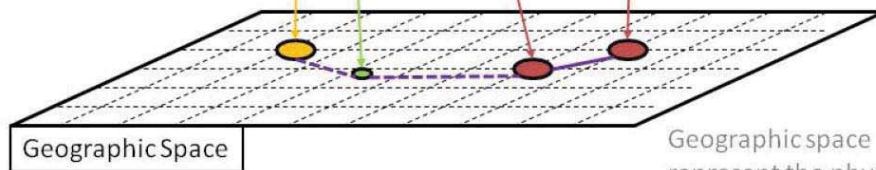


Figure 1: Diagram representing the main mechanics of the ENTRABMS mode

4.0 CONCLUSION

Our study uses an agent-based model to study new venture formation. The primary variables of interest that are examined include the levels of new venture formation across different locations in the geographic space and the rate of survival of new ventures in this space. The results measured by these variables will display variable results from the dynamics of the system and its initial conditions, which are established with several systemic parameters. These include the number of competitors, distribution of customer demand, presence of institutional agents, and the rate of entry.

The anticipated results from this simulation study will reinforce the existing premise in the literature that formal networks are important to the formation of new ventures,

but also introduce evidence that the structure of those networks may have dramatically different outcomes based on the nature of the institutional environment the venture resides in. New ventures should seek out and develop the 'right' networks for their situation and location. What 'right' looks like depends on nature of the new venture and its knowledge appropriation process, and the type of environment the entrepreneur is operating in.

In addition, this study will demonstrate how simulation may be used to evaluate the presence of competing theories of behavior on heterogeneous agents in a landscape model. This work contributes to the scholarly direction of entrepreneurship research while advancing the agent-based modeling methodology in the social sciences domain.

REFERENCES

- [1] Audretsch, D. (2002); *Entrepreneurship: A survey of the literature*. London: European Comm., Enterprise Directorate General.
- [2] McKelvey, B. (2004); "Toward a Complexity Science of Entrepreneurship." *Journal of Business Venturing*, 19(3): 313.
- [3] Carayannis, E. G., Provance, M., and Givens, N. (In press, available online 2011); "Knowledge Arbitrage, Serendipity, and Knowledge Formality: Their Effects on Sustainable Entrepreneurial Action in Regions." *IEEE Transactions on Engineering Management*.
- [4] Provance, M. (2010); *Tying It Together: Three Essays on the Roles of Knowledge, Entrepreneurial Action, and Institutions in the Formation of New Ventures*; Proquest UMI.
- [5] Bruton, G. D., Ahlstrom, D., and Li, H.-L. (2010); "Institutional Theory and Entrepreneurship: Where Are We Now and Where Do We Need to Move in the Future?" *Entrepreneurship Theory and Practice*, 34(3): 421-440.
- [6] Boulding, K. (1981 [1978]); *Ecodynamics*. Beverly Hills, CA: SAGE Publications.
- [7] Scott, R., and Meyer, J. (1994); *Institutional Environments and Organizations*. Thousand Oaks, CA: SAGE Publications.
- [8] Fligstein, N. (2001); *The Architecture of Markets*. Princeton, NJ: Princeton University Press.
- [9] North, D. C. (1991); "Institutions." *The Journal of Economic Perspectives*, 5(1).
- [10] Hamilton, W. R. (1852); *Discussions on philosophy and literature, education and university reform*. London.
- [11] Rodriguez-Fernández, J.L. (1999); "Ockham's razor" *Endeavour* 23(3); 121-125.
- [12] Pidd, M. (1996); *Tools for Thinking: Modelling in Management Science*; Wiley.
- [13] North, M.J. and Macal, C.M. (2007); *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*; Oxford University Press.
- [14] Miller, J.H. and Page, S.E. (2007); *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*; Princeton, NJ: Princeton University Press.
- [15] Wolfram, S. (2002); *A New Kind of Science*; Champaign, IL: Wolfram Media.
- [16] Home, G. and Meyer, T. (2010); "Data Farming and Defense Applications" *Proceedings of the MODSIM World 2010 Conference*.
- [17] Bitinas, E., Henscheid, Z., and Middleton, D. (2005); "Pythagoras: An Agent-Based Environment"; Working paper; Agent-Based Models and Other Analytical Tools in Support of Stability Operations Workshop, Military Operations Research Society, Science Applications International Corporation, McLean, Virginia.
- [18] North, M.J., Howe, T.R., Collier, N.T., and Vos, J.R. (2007); "A Declarative Model Assembly Infrastructure for Verification and Validation" in Takahashi, S., Sallach, D.L., and Rouchier, J. (eds.), *Advancing Social Simulation: The First World Congress*; Germany: Springer.

4.0 EDUCATION TRACK

Track Chair – Vincent Betro, University of Tennessee SimCenter



4.1	Overcoming the Critical Shortage of STEM – Prepared Secondary Students Through Modeling and Simulation	306
4.2	Desktop Modeling and Simulation: Parsimonious, Yet Effective Discrete-Event Simulation Analysis	312
4.3	The Supply Chain Game: Serious Gaming in Supply Chain Management Curricula	320
4.4	Effects of a Pedagogical Agent’s Emotional Expressiveness on Learner Perceptions	328

4.1 Overcoming the Critical Shortage of STEM – Prepared Secondary Students Through Modeling and Simulation

Overcoming the Critical Shortage of STEM – Prepared Secondary Students Through Modeling and Simulation

Tom Spencer and Brandon Berry – Pre-Engineering Technology Academy Instructors, Grassfield High School, Chesapeake, VA

thomas.spencer@cpschools.com brandon.berry@cpschools.com

Abstract:

In developing understanding of technological systems – modeling and simulation tools aid significantly in the learning and visualization processes. In design courses we sketch, extrude, shape, refine and animate with virtual tools in 3D. Final designs are built using a 3D printer. Aspiring architects create spaces with realistic materials and lighting schemes rendered on model surfaces to create breath-taking walk-throughs of virtual spaces. Digital Electronics students design systems that address real-world needs. Designs are simulated in virtual circuits to provide proof of concept before physical construction. This vastly increases students' ability to design and build complex systems. We find students using modeling and simulation in the learning process, assimilate information at a much faster pace and engage more deeply in learning.

As Pre-Engineering educators within the Career and Technical Education program at our school division's Technology Academy our task is to help learners in their quest to develop deep understanding of complex technological systems in a variety of engineering disciplines. Today's young learners have vast opportunities to learn with tools that many of us only dreamed about a decade or so ago when we were engaged in engineering and other technical studies. Today's learner paints with a virtual brush - scenes that can aid significantly in the learning and visualization processes. Modeling and simulation systems have become the new standard tool set in the technical classroom [1-5].

Modeling and simulation systems are now applied as feedback loops in the learning environment. Much of the study of behavior change through the use of feedback loops can be attributed to Stanford Psychologist Alfred Bandura. "Drawing on several education experiments involving children, Bandura observed that giving individuals a clear goal and a means to evaluate their progress toward that goal greatly increased the likelihood that they would achieve it." [6].

As a young sailor racing small-scale sail boats I was told by an experienced competitor that I needed to develop "a really good sense of speed." This meant that I needed to be able to

perceive the associated increase or decrease in the boat's speed as I made a change. I could then optimize these small changes to outperform other competitors. This little story illustrates what we are attempting to do with modeling and simulation systems employed in pre-engineering and Career and Technical Education. Having learners reach the "Evaluation" stage in Bloom's taxonomy – is the the holy grail for many educational initiatives. Students would then be able to gauge their progress and correct their course accordingly. However, we restrict this goal to the competencies which our learners are tasked with – which should be achievable goals on an individual basis. The larger task of educating fully-prepared professional engineers will be left to the university environment. [7-8].

Students are applying modeling and simulation systems as feedback loops that inform their learning process. They construct and test an element of their design and the modeling and simulation system allows them to explore its attributes and functionality in-process prior to moving forward. In this feedback loop – the student is not required to consult an outside source – such as the instructor or another student to test basic relationships and design ideas. The feedback offered is nearly immediate and very dynamic. The modeling and simulation system will allow for a rich

exploration of the learner's creation which is used to improve design ideas in-process. This has the effect of deepening and enriching interactions between learners, their peers and instructors due to the fact that they are often interacting about larger more systematic issues that are central to the functionality of the design [9-10].

Carnegie Mellon University implemented an instructional initiative which deployed custom-developed software designed to give individual immediate feedback to support learners in its Open Learning initiative which consists primarily of mathematics and science courses. The goal of this grant-funded project was to improve the learning experience for learners who did not have access to campus-based lectures or labs through software that would give learners individual feedback. When compared to a control group of campus-based learners who had the same course-work without access to lectures and labs – there was no significant difference found between the two groups performance on standardized tests. Rather than pursue the thought of eliminating teaching positions in favor of software – Carnegie Mellon is now exploring deployment of this software in live, on-campus courses to allow instruction to accelerate - better-utilizing the expert abilities of their instructors by allowing them to cover topics in more depth and help learners to see their real-world applications[11].

In mechanical design courses learners sketch their ideas in two dimensional virtual space using software such as Autodesk Inventor. They then extrude their ideas into three dimensional form – applying concepts from geometry and trigonometry continuing to mold and shape with virtual tools creating objects that exist only in digital form. These forms can be freely manipulated within the software – allowing learners to see any view of the object they desire to aid in the visualization of their creation. Learners can animate their creation which will behave based-on physical constraints applied in the model before it is built in real life. Finally learners can send the model file - previously only a construct in software - to a 3D printer – creating in three dimensions a part that now exists in the real

world[12].

Mechanical Design – Catapult

- Task was to create a 3d model of a given catapult and modify it to improve its design.
- Students did this by changing the power source from a tension system to one that uses a counterweight.
- Beginning of video is a loop showing how the catapult would react upon firing in real life using the force of gravity on the 1 lb counterweight.
- Next view is rotated to show different angles of the device.
- Lastly video zooms in to the browser to show all of the individual parts that have been created by the students

Mechanical Design – Lab Support Stand (Scissor Jack Mechanism)

- Task - create a 3D model of the given lab support stand and properly constrain all parts to simulate its motion
- Beginning of video loops through multiple turns of the Acme thread screw to fully extend – then collapse the scissor jack mechanism of the lab support stand

Aspiring young architects and civil engineers create models in software that allow them to effectively simulate the professional tasks required of professional practitioners. Again the learners are using some of the most practical applications you might imagine from all of their prior mathematics instruction – particularly geometry and trigonometry. Architects and civil engineers utilize modeling and simulation software as a project data management system and as a coordination tool for their interaction. Both disciplines also use the system to develop presentations of their work to clients, governing bodies and other interested parties.

Learners in Architecture courses create spaces with realistic materials, many flexible elevation views realistic and lighting schemes rendered

upon the model surfaces to create breath-taking walk-throughs of virtual spaces that can be shown to potential clients or users of a building. The modeling and simulation package, Autodesk Revit includes tools to offer clients a variety of design options from which to choose – enhancing the possibility that a potential client can have the opportunity to choose a satisfactory design solution. The system also allows the user to perform site opportunities analysis – optimizing the siting of the structure to allow for day-lighting, wind protection and opportunities for pleasing views of the surrounding landscape. They can even model the natural light in the various seasons of the year for a particular location. The internal structures within wall systems are still 2D elements that require significant detailing due to the variable construction practices that are used in different geographical areas of the country. This produces a wide variety of choices that have to be made without automation – an opportunity to explain the challenges of professional practice[13-14].

Civil projects have an even greater variability as much of the information about a project, it's benchmark, elevation, topographical data and property boundaries must all be acquired and effectively integrated into the project files with a high degree of accuracy. Some of this work is generated from 2D drawings and other data references this 2D data as the project is transformed into a 3D project file. The opportunities for error are many - admonishing learners to develop excellent habits when they are working with project data. Again, Autodesk Revit offers the Civil Engineer a tool that is capable in the 3D realm while allowing traditional 2D format site data to be easily integrated[15].

Architectural / Civil Design -

- Task – design a single-family dwelling complete with interior and exterior detailing with a “walk-through” visualization for potential clients
- Problem is addressed by using Autodesk Revit for the site plan, floor plans and elevations - followed by application of realistic materials and lighting appropriate to the site in

Autodesk 3D Studio Max

- Learners get experience simulating appropriate spatial relationships in three dimensions and receive rich feedback as they see design choices illuminated in the walk-through of their design.

Digital Electronics learners design systems that address real-world needs – providing the ability to sense the state of systems and respond accordingly. We teach that digital electronics are only of use if they can effectively interface with the real world – which is analog in nature. Learners use National Instruments' MultiSIM to simulate their designs in virtual circuits – providing a proof of concept before investing the time in building the circuits in real life. This vastly increases learners' ability to confidently design and build challenging systems with many input variables and many output states. Trial and error testing is easily accomplished in a safe and flexible test environment. Learners may actually learn more due to the ease with which they can play “what-if” with the simulation. In real-world circuitry – some of the what-if scenarios would produce component damage that would create the need for a potentially time-intensive and frustrating trouble-shooting process. This time can be better-utilized learning how to implement more complex designs or for adaptation to real-world situations.

Electronics Engineering / Design – 60 Second Timer

- Task - construct a 60 second timer utilizing D Flip-Flops and 7 Segment Display Drivers
- Problem is addressed by configuring the Flip-Flop stages as decade counters
- First stage (one's place) counts 0 to 9 then carries over to the second stage (ten's place)
- The second stage has AOI reset logic that resets the counter when the counter sees a 7 – however, the user only sees a 6 on the display in the tens position – as the reset logic acts within milliseconds.
- The user may manually reset/restart the counter by pressing a push button

- normally open switch (PBNO)
- The 7 Segment Display drivers decode the binary count and Drive a one's place and a ten's place (SSD)

In summary, we would suggest that contemporary learners in technical coursework where modeling and simulation software is available can best be served by integrating these powerful tools into instructional practice. There is a powerful feedback loop that forms between the learner's experimentation with design ideas and the near-immediate feedback from the modeling and simulation system that informs the the design process. Ideas for

design of new forms that have beneficial characteristics can be honed prior to prototype creation and eventual data-acquisition from a physical model. This reduction in the number of physical models that must be constructed and tested while learning the basics of the applied discipline allows the learner to acquire new knowledge at a faster pace. Interaction between the learner, peers and the instructor may be improved by focusing the interaction required on larger, more systematic concepts – deepening the learning process for all parties[16.]

References

[1]. Title:Meeting the challenges of STEM education: Teaching Modeling and Simulation with real-world applications

Authors:Banks, Catherine M. (1); Sokolowski, John A. (1)

Author affiliation:(1) Virginia Modeling, Analysis and Simulation Center, Old Dominion University, United States

Corresponding author:Banks, C. M.
(cmbanks@odu.edu)

[2]. Source title:Spring Simulation Multiconference 2010, SpringSim'10

Abbreviated source title:Spring Simul. Multiconference, SpringSim

Monograph title:Spring Simulation Multiconference 2010, SpringSim'10

Issue date:2010

Publication year:2010

Article number:58

[3]. The role of digital modeling and simulation in power engineering education

Authors:Kozunovic, M. (1); Abur, A. (1); Garng Huang (1); Bose, A.; Tomsovic, K.

Author affiliation:(1) Dept. of Electr. Eng., Texas A&M Univ., College Station, TX, USA

Source title:IEEE Transactions on Power Systems

Abbreviated source title:IEEE Trans. Power Syst. (USA)

Volume:19

Issue:1

Publication date:Feb. 2004

Pages:64-72

[4]. Modeling and simulation in high school education - two European examples

Authors:Herper, H. (1); Stahl, I.

Author affiliation:(1) Inst. for Simulation & Graphics, Otto-von-Guericke Univ., Magdeburg, Germany

Source:Proceedings of the 2003 Winter Simulation Conference (IEEE Cat. No.03CH7499)

Publication date:2003

Part number:2

Part number:2
Pages:1973-81 vol.2

[5]. Simulation in education: modeling for the future
Authors:Sowers, J.L. (1); Oren, T.; Ziegler, B.; Houska, R.V.; Kopka, R.
Author affiliation:(1) Rutgers, State Univ., Camden, NJ, USA
Source:1983 Winter Simulation Conference Proceedings
Publication date:1983
Pages:259-60

[6]. Harnessing the Power of Feedback Loops
Authors: Thomas Goetz
Author affiliation: (1) Wired Magazine
Publication Date: July 2011
Source: http://www.wired.com/magazine/2011/06/ff_feedbackloop/

[7]. 2011/2012 Competency-Based Task/Competency Lists for Career and Technical Education Courses (Multiple Courses)
Authors: Virginia Department of Education
Author affiliation: (1) Virginia Department of Education
Issue: Academic Years 2011, 2012
Source:
http://www.cteresource.org/verso2/results?program_area=technology_education&document_type=all&course_codes=&text=

[8]. An Ontology and a Software Framework for Competency Modeling and Management.
Authors: Paquette, Gilbert.
Source: Journal of Educational Technology & Society, Jul2007, Vol. 10 Issue 3, p1-21

[9]. Interactive strategies for developing intuitive knowledge as basis for simulation modeling education
Authors:Atolagbe, T.A. (1); Hlupic, V. (1); Taylor, S.J.E. (1); Paul, R.J. (1)
Author affiliation:(1) Defpt. of Inf. Syst. & Comput., Brunel Univ., Uxbridge, UK
Source:Proceedings of the 1997 Winter Simulation Conference
Publication date:1997
Pages:1394-402

[10]. A review of modeling and simulation in education and training
Authors:Wildberger, A.M. (1)
Author affiliation:(1) Naval Ordnance Lab., Silver Springs, MD, USA
Source:1974 Winter Simulation Conference, Vol.II
Publication date:1974
Pages:735-6

[11]. Hybrid Education 2.0
Author: Steve Kolowich
Author Affiliation: Insidehighered.com
Publication Date: December 28, 2009
Source: <http://www.insidehighered.com/news/2009/12/28/carnegie>

- [12]. A Study of the Efficacy of Project-based Learning Integrated with Computer based Simulation – STELLA.
Authors: Eskrootchi, Rogheyeh; Oskrochi1, G. Reza.
Publication: Journal of Educational Technology & Society, Jan2010, Vol. 13 Issue 1, p236-245
- [13]. Geotechnical Physical Modeling for Education: Learning Theory Approach. By: Wartman, Joseph. Journal of Professional Issues in Engineering Education & Practice, Oct2006, Vol. 132 Issue 4, p288-296
- [14]. Students' reasoning during modeling in an inquiry learning environment. By: Löhner, Simone; van Joolingen, Wouter R.; Savelsbergh, Elwin R.; van Hout-Wolters, Bernadette. Computers in Human Behavior, May2005, Vol. 21 Issue 3, p441-461
- [15]. Spatial Cognition Support for Exploring the Design Mechanics of Building Structures. By: Rudy, Margit; Richard Hauck. Journal of Interactive Learning Research, 2008, Vol. 19 Issue 3, p509-530
- [16]. Digital Modeling in Design Foundation Coursework: An Exploratory Study off the Effectiveness of Conceptual Design Software. By: Guidera, Stan; MacPherson, D. Scot. Journal of Technology Studies, Spring2009, Vol. 34 Issue 1, p55-66

4.2 Desktop Modeling and Simulation: Parsimonious, Yet Effective Discrete-Event Simulation Analysis

Desktop Modeling and Simulation: Parsimonious, Yet Effective Discrete-Event Simulation Analysis

James R. Bradley
Mason School of Business
College of William and Mary
james.bradley@mason.wm.edu

This paper evaluates how quickly students can be trained to construct useful discrete-event simulation models using Excel. The typical supply chain used by many large national retailers is described, and an Excel-based simulation model is constructed of it. The set of programming and simulation skills required for development of that model are then determined: we conclude that six hours of training are required to teach the skills to MBA students. The simulation presented here contains all fundamental functionality of a simulation model, and so our result holds for any discrete-event simulation model. We argue, therefore, that industry workers with the same technical skill set as students having completed one year in an MBA program can be quickly trained to construct simulation models. This result gives credence to the efficacy of *Desktop Modeling and Simulation* whereby simulation analyses can be quickly developed, run, and analyzed with widely available software, namely Excel.

1.0 INTRODUCTION

Simulation analyses in business typically use one of two simulation methodologies, Monte Carlo simulation or discrete-event simulation. These analyses can be conducted with many different types of software platforms, including (1) commercial discrete-event and Monte Carlo software packages, and (2) custom programs written in Visual Basic, C#, Java, or other similar programming language. Advantages of the first approach include avoiding the need to program the functionality that comes with a commercially available product, such as the generation of random variables and, for discrete-event simulation, graphic user interfaces that make constructing a model more intuitive. One disadvantage of using an existing simulation program is their cost, which can be thousands of dollars. Advantages of authoring one's own simulation program is that it is likely to run much faster than commercial packages and custom features can be created that are not available in packaged software. Disadvantages of custom simulation models are acquiring the requisite programming skills and the time required for programming. Also, programming animation, if it is desired, can add considerably to development time.

In this paper, we consider an alternative to these two "endpoints" on a continuum of

software implementations of simulation. Specifically, we consider developing discrete-event software analyses in Excel, supplemented by simple Visual Basic for Applications (VBA) code. (While we focus on discrete-event simulation in this paper, our conclusion is likely to be applicable also to Monte Carlo simulation as is investigated by Guerrero (2010b).) We show that constructing a complex simulation with this ubiquitous software package can be accomplished with only minimal training in VBA. This approach mitigates many of the disadvantages of the two endpoint alternatives. First, Excel is inexpensive, since virtually everybody has access to it on their personal computer. Second, although the VBA knowledge required is minimal compared to that required to generate custom programs, many users would find Excel/VBA to be not much more difficult, if at all, than learning the graphical user interface for simulation software platforms. Third, while animation of processes is sometimes useful, most often business decisions are driven the simulation data rather than animation. Animation is, by the way, possible in Excel although it requires additional VBA programming knowledge.

This paper supports the foregoing assertions about discrete-event simulations by showing an example simulation of a supply chain where decisions are made

regarding replenishment of goods from an overseas source. This model is coded in the simplest manner possible in order to determine the minimum possible training to enable a user to construct a valid simulation. Our focus, then, is on how the simulation is implemented rather than on analyzing the simulation results. We find that Excel can be used to quickly construct such models, which provides a basis of support for a trend in simulation, which has been called "Desktop Modeling and Simulation" by Guerrero (2010b). The hallmarks of desktop modeling and simulation are that small models can be quickly constructed in business by individuals with a minimum of training to make effective decisions on possibly complex business scenarios.

2.0 BODY

2.1 Simulation Context

The context for the sample discrete-event model used in this paper is a global supply chain, which is the type that might be operated by a large national retailer (e.g., Wal-Mart, Target). The supply chain has five links (see Figure 1):

1. Overseas source of goods
2. Crossdocking facility
3. Import distribution Center (IDC)
4. Regional Distribution Centers (RDCs)
5. Retail Stores

The first two links are connected, primarily, by oceangoing container ships and logistics at the originating port. The remaining links are in the US and are linked in this simulation model by truck transportation. (In other contexts, rail or barge transport might be involved.)

Quickly described, each of these links serves the following purposes: the overseas source can be a manufacturer or distributor of goods that might possibly go through a consolidation step before arriving at the port

of origin; the Crossdocking facility unloads the incoming containers, which may contain only a small number of stock keeping units (SKUs) and allocates the large quantity of each SKU in each container into smaller quantities in many trailers destined for RDCs or retail stores; the remainder of the incoming inventory not allocated to RDCs is sent to the Import Distribution Center (IDC) for storage until such time as it is needed at the RDCs; and the Retail Stores which, obviously, make the goods available to customers.

Typically in such a supply chain, ships arrive at a port close to the IDC from each origin every week or so with a particular SKU whereas, in the intervening time, the inventories temporarily stored in the IDC can be used to replenish the RDCs and retail stores that experience heavy demand.

The purpose of the supply chain simulation is to determine decisions about:

1. Quantities of imported goods to immediately allocate to RDCs
2. Daily replenishment quantities from the IDC to the RDCs

Retaining inventory at the IDC until the need arises at the RDCs rather than sending it immediately to the RDCs is called *postponement*. This is an emerging supply chain management tactic that can reduce inventory levels. Alternately, if all the incoming inventories are immediately sent on to RDCs, then a company would be vulnerable to fluctuating demand; if demand at one RDC were higher than expected, which will invariably happen, then that RDC would run out of stock while inventory would be left over at other RDCs where demand is low. This situation is expensive to resolve (i.e., transshipment among RDCs can be used), or infeasible due to financial considerations. Thus, inventories are high (leftover inventory at some RDCs) while service to customer is low (at some RDCs that experience high demand). Waiting to observe actual demand before allocating

goods allows better placement of inventory, which can reduce investment in inventory and improve availability of goods to customers. Just how one should implement an appropriate postponement strategy, however, is difficult to analyze. Analysis with calculus to determine an optimal decision policy is intractable, and so simulation offers a viable means of studying such a system.

2.2 Simulation Model Description

The main worksheet in the simulation workbook is shown in Figure 2. Note that the sheet is designed in accordance with “spreadsheet feng shui” as advocated by Guerrero (2010a). Specifically, color coding denotes regions of the worksheet that serve different purposes (problem parameters, ordering decision policies, day-to-day orders resulting from ordering policies, daily inventory status at IDC and RDCs, etc.), which makes the simulation more readily understandable.

The two parameters of the ordering decision policy are two target service levels for the weekly replenishment from arriving imports and the daily replenishment from the IDC. The 60% service level in Figure 2 indicates that when an imported shipment arrives RDCs each receive enough inventory such that in 60% of the weeks they would not stock out. The daily replenishment target service level of 97% indicates that on a daily basis each RDC receives a shipment quantity that is sufficient such that 97% of the days they would not stock out. With both weekly and daily replenishment it is possible that either the incoming imported shipment or the inventory at the IDC (respectively) is sometimes not sufficient to reach these service level targets, in which case an alternate allocation scheme is used. This service policy is a modification of a well known optimal policy in the inventory management literature (i.e., the optimal solution to a multiple-news vendor model with constrained supply).

The crux of the decision is to determine how much of the incoming imported goods to allocate immediately to RDCs: transportation cost can be saved if more inventory is sent immediately to the RDCs whereas, if too much inventory is sent, it is likely to be sent to the “wrong” RDC such that one RDC will run out of inventory while another RDC has leftover inventory.

Figure 3 shows how the resulting fill rate on the y-axis (the percentage of customers immediately satisfied) and the average weekly cost of inventory and shipping (x-axis) changes as the service level from inventory immediately allocated from imports increases from 10% to 80%. The simulation reveals that the fill rate, at first, increases as weekly target service level from imports increases, but fill rate decreases as the weekly service level increases past 40%. This is due to, as described above, the increasing frequency of one RDC having leftover inventory at the end of a week while other RDCs have run out of inventory.

2.3 Simulation Model Structure

The key components of a basic discrete event simulation algorithm are:

1. Generating random variables that represent uncertain events
2. Keeping track of the state of the system as the simulation progresses
3. Saving simulation data for later analysis
4. Displaying simulation results
5. Controlling how long the simulation is run

This simulation has been constructed with the intent of minimizing (i) the amount of knowledge one must have regarding the general methodology of simulation, and (ii) expertise in spreadsheet construction and VBA programming, which motivated the particular programming tack that we describe below. Spreadsheet simulation functionality can be implemented in many

different ways, for example, within the spreadsheet, in the VBA “behind” the spreadsheet, in an Excel add-in, or in another programming language which is compiled and linked to the spreadsheet. Our goal of minimizing the training required to implement simulation motivates us to avoid programmatic approaches where possible in favor of using the Excel spreadsheet itself because many more people are familiar with Excel than with computer programming languages.

Accordingly, rather than generating random variables with an algorithm coded in VBA or some other programming language, we use the spreadsheet in the simulation model. In the global supply chain spreadsheet, we model daily demand at the RDCs as being normally distributed. Thus, we use the RAND() spreadsheet function (which generates a random variable from the continuous uniform distribution) together with the NORMINV() spreadsheet function to “draw” random variables using the inverse probability function. In addition, we use the MAX() function to ensure that demand is not less than zero (which makes no sense in this context). Using a distribution’s inverse is not always feasible. For example, for discrete probability distributions where Excel does not offer an inverse function, the target cumulative probability distribution can be computed in a range within a worksheet such that random variables can be generated using the RAND() and VLOOKUP() spreadsheet functions. The probability function range would include at a minimum the possible values of the random variable (in ascending order) and a column for the corresponding cumulative probability function. VLOOKUP() can then be used to identify the appropriate random variable value which could be found in the row where the cumulative probability first exceeds the RAND() value. A similar approach is possible using the MATCH() spreadsheet function. Implementing inverse probability functions in this manner within the spreadsheet is likely easier to teach to

programming neophytes than a programmatic approach.

The state of the system at any point of time consists of the inventory levels at the IDC and at each RDC. Showing inventory levels visually on the worksheet is beneficial so that users can observe what is happening, but it requires VBA programming. Perhaps the simplest approach is to designate one cell in the simulation worksheet to contain each inventory quantity and then use the Range.Offset.Value VBA statement to update inventory levels at the end of each day. Inventory at the end of the day (the new value in the cell) is the old value in the cell plus the quantity of inventory received, minus the outgoing inventory. (Note that a simplistic approach using a worksheet function to compute this value without VBA results in a circular reference.) For example, the following statement is used to update the inventory count at each RDC based on the amount of inventory shipped from the IDC to the RDCs, where *shipQ* is a VBA variable that denotes the quantity shipped from the IDC to the RDC in question:

```
.Range("simInvRDC").Offset(RDC, 0).Value  
= .Range("simInvRDC").Offset(RDC,  
0).Value + shipQ
```

A similar statement is used later in the program to reduce the RDC inventory due to goods that are delivered to retail stores and taken out of inventory. Note that the VBA code makes extensive use of named ranges, such as “simInvRDC”, to make VBA code more readable. If-then VBA statements are also required to account for inventory, as well as understanding the basic arithmetic operators in VBA (+, -, *, /).

In the case of the example simulation, storing inventory level history at each of the DCs is useful for subsequent analysis. Thus, we keep track of the IDC and RDC inventory in a “Results” worksheet. These data can then be graphed to better help understand the functioning of the supply chain over time under the proposed order

policy. Storing these data requires VBA programming, albeit perhaps the simplest statement to understand, namely some form of the aforementioned `Range.Offset.Value` command, which can be used to copy the current inventory levels from the simulation worksheet onto the appropriate row and column of the “Results” worksheet. When the range for the simulation output can be identified in advance, a graph can be constructed beforehand to display results automatically when the simulation finishes.

Controlling the number of weeks over which the simulation is run requires VBA also. In this case, we have used a For-Next loop in VBA to run the simulation for the predetermined number of weeks indicated on the main simulation worksheet.

Note that the simulation worksheet must be recalculated each day in order to refresh (generate) new demand at each RDC. This requires either the `Worksheet.Calculate` or the `Range.Calculate` VBA function because, as we describe below, it is normally wise to turn off the automatic workbook recalculation.

While the above functionality arguably constitutes the basic implementation of a simulation analysis, additionally, other knowledge is advantageous to improve the performance of the simulation. Specifically, to increase the speed at which the simulation runs, it is beneficial in Excel to suspend the graphical updating of the spreadsheet and, also, to suspend the recalculation of the spreadsheet, which in automatic mode occurs whenever the value in any cell changes. Otherwise, the simulation can be very slow. To disable these features, one must use these commands:

- `Application.Calculation = xlCalculationManual`
- `Application.ScreenUpdating = False`

It is good practice to turn screen updating and recalculation back on at the end of the simulation with these statements:

- `Application.Calculation = xlCalculationAutomatic`
- `Application.ScreenUpdating = True`

Also, while not strictly necessary, we have declared VBA variables in the simulation, which helps to avoid programming errors.

2.4 Required Ancillary Skills

In order to draw appropriate conclusions from a simulation analysis, one must understand hypothesis testing, or alternative statistical methods that can discern the significance of differences in sample statistics generated by simulation analyses. In the target population discussed earlier, MBA students have already been exposed to hypothesis testing by the time they encounter a supply chain management course, so this is a moot point. This topic would need to be covered for some other populations, however. In addition, for MBA students, and possibly all students, must be made aware of the Law of Large Numbers and the Central Limit Theorem in order to understand the necessity of sometimes increasing the runtime of simulations in order to obtain definitive results.

2.5 The “Next” Simulation Skills

We mention one other programming technique that is useful when the data generated by simulation analyses is large. That is, the data from simulation studies can be stored in a database rather than in a worksheet as described above. This is beneficial when the data are sizeable. VBA offers connectivity to Access, SQL Server, and other databases. One should take care, however, to avoid using Access when the data cause the database to exceed roughly 2GB in size. At that point Access fails and, in the author’s experience, gives no clear error message to indicate the specifics of the error.

3.0 DISCUSSION

A summary of the knowledge required to implement the simulation model as described in Section 2.3 is as follows:

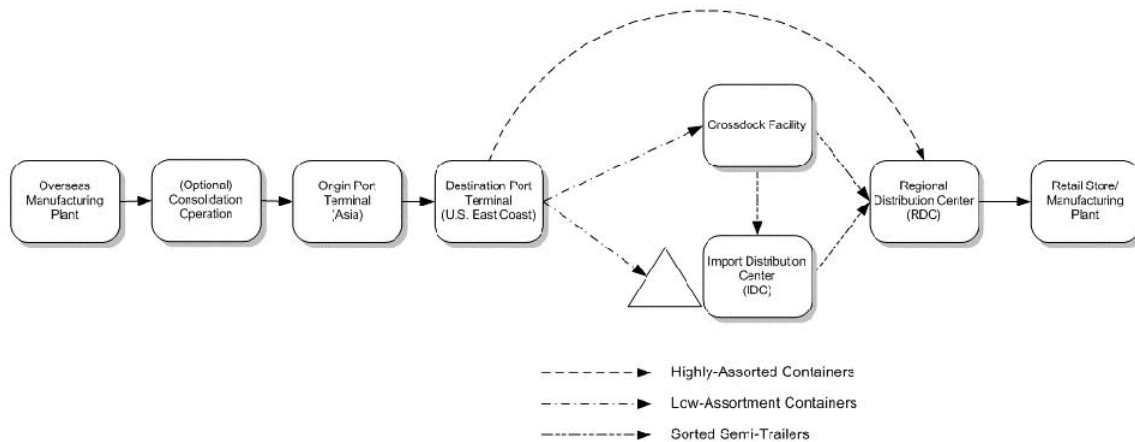


Figure 1: Diagram of Example Global Supply Chain

Global Supply Chain Distribution Simulation Analysis v5.xlsx - Microsoft Excel									
Ordering Decision Policies									
Postponement Analysis for Global Distribution Supply Chain									
Simulation Parameters				Cost Parameters					
Quantity per Container	5,000			Transport: Port to Crossdock	\$ 200.00	Container Crossdocking	\$ 400.00		
Weeks to Simulate	100			Transport: Port to IDC	\$ 250.00	IDC Handling cost (per case)	\$ 0.25		
				Transport: Crossdock to IDC	\$ 150.00				
				Transport: IDC to RDC (per case)	\$ 0.50				
				Transport: Crossdock to RDC	\$ 300.00				
Ordering Decision Policies				James Bradley: Decision variable					
Weekly Service Level	60%								
Daily Service Level	97%								
Inventory allocation:	FCFS								
This Week				Simulate					
Week									
Containers	2								
This Week's Shipment	9,466								
Cumulative Cost									
% In Stock									
Average Weekly Cost	#DIV/0!								
								IDC Inventory	110
Demand Characteristics			Ordering Decisions			Shipments		Status	
Distribution Center	Mean Daily Demand	Standard Deviation of Daily Demand	Weekly Order-Up-To Point	Daily Order-Up-To Point	Daily Demand	From Crossdock	From IDC	Inventory/Backorders	Backorders
1	142.9	28.6	1019	197	128	965	47	71	0
2	146.4	73.7	1074	284	43	885	56	98	0
3	285.7	85.7	2057	447	261	1905	0	475	0
4	214.3	85.7	1557	375	456	1405	0	263	0
5	321.4	96.4	2815	538	214	2126	236	237	0
6	139.3	69.6	1022	270	31	982	268	143	0
7	178.6	71.4	1278	313	175	1102	145	130	0
Total Weekly Demand	10000.0	530.0				9400	752		0

Figure 2: Main Worksheet of Simulation Workbook

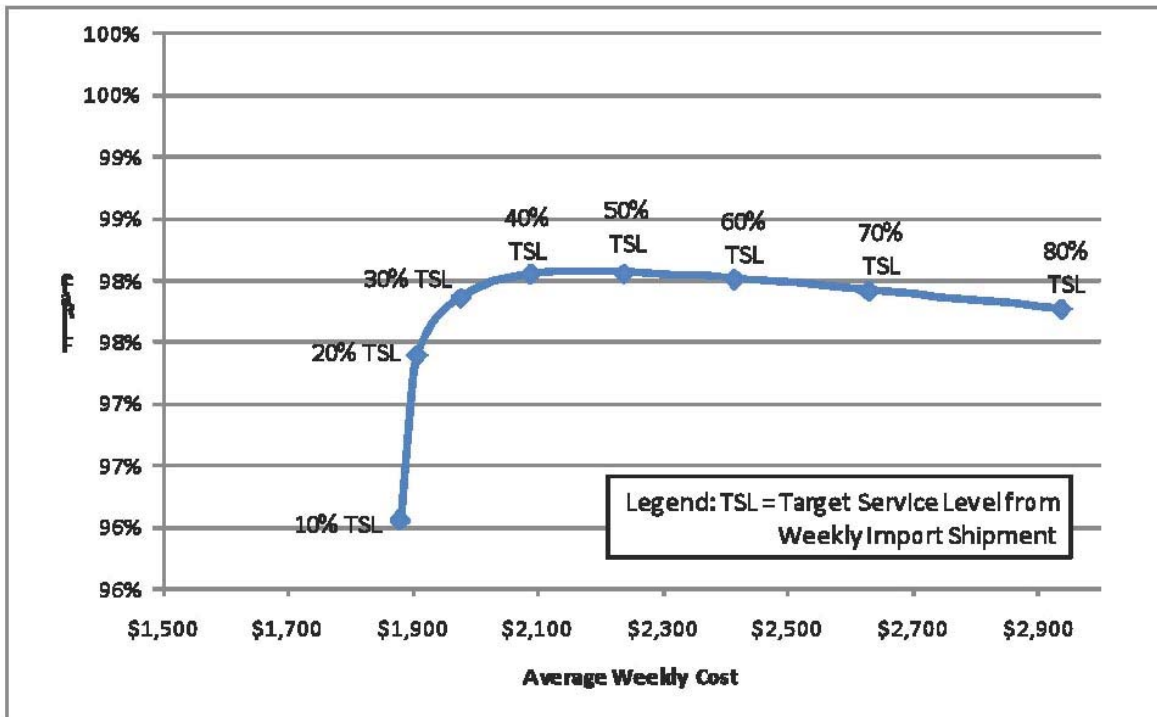


Figure 3: Simulation Results

1. General programming knowledge
 - a. Basic understanding of a computer program as a set of steps that are automatically executed
2. Excel spreadsheet knowledge
 - a. Understanding of manual vs. automatic recalculation
 - b. Named ranges
 - c. CommandButton Form Object
3. VBA programming knowledge
 - a. Programming statements
 - i. Declaring variables
 - ii. Basic arithmetic operators
 - iii. Worksheet.Range.Offset.Value
 - iv. Application.ScreenUpdating
 - v. Application.Calculation
 - vi. Range.Calculate
 - vii. For-Next loops
 - viii. If-Then
 - ix. With statement
 - b. VBA concepts
 - i. Difference between changing a cell value versus the value of a VBA variable
 - ii. Integer/Double variable types
 - iii. Basic Debugging and break points
4. Statistical knowledge
 - a. Hypothesis testing
 - b. Law of Large Numbers
 - c. Central Limit Theorem

The VBA With statement in the list above has not yet been mentioned. While this statement is not necessary to implement the example simulation, it does render the code more readable and benefits the students in later programming ventures. Additionally, we have not discussed basic debugging skills, such as setting break points which, while not absolutely necessary, are useful.

Instruction in VBA programming has been included in several courses at the Mason School of Business at the College of William and Mary. From this experience the author estimates that, in either the undergraduate or MBA programs that the Excel and VBA skills in the list above can be taught in four 1.5 hour class sessions, with follow-on

homework assignments. People with less comfort or experience with Excel may require more sessions.

If one had a goal of educating people in industry to do a simulation from which valid conclusions were drawn, we should expect the foundation of many people in industry to be comparable with the skills of a person entering an MBA program. In that case, we should expect the training requirements for such a person to be commensurate with the training required for MBA students, with the exception that some of the statistical training that an MBA student receives may be required. Thus, one could draw the conclusion that a person from industry could be trained with the necessary skills to complete a valid simulation study in a reasonably short time.

4.0 CONCLUSION(S)

This paper documents a discrete-event simulation program which could be used by companies to effectively inform an important supply chain decision. The analysis of the skills required to construct that program, which are also sufficient for other similar simulation models, suggests that simple, yet effective simulation skills can be taught to a many workers in industry. This is but one example of what one might call "Desktop Modeling and Simulation."

5.0 REFERENCES

Guerrero, Héctor H. (2010a), *Excel Data Analysis: Modeling and Simulation*, Springer, New York.

Guerrero, Héctor H. (2010b). "Essentials for Monte Carlo Simulation with Excel," submission to MODSIM World 2010 Conference.

6.0 ACKNOWLEDGMENT(S)

The author thanks Héctor H. Guerrero, who has been a wonderful colleague and from whom the author has benefitted enormously in discussions as they have developed the notion of Desktop Modeling and Simulation.

4.3 The Supply Chain Game: Serious Gaming in Supply Chain Management Curricula

The Supply Chain Game: Serious Gaming in Supply Chain Management Curricula

James R. Bradley
Mason School of Business
College of William and Mary
james.bradley@mason.wm.edu

Discrete-event simulation has long been used for research and evaluation of decision alternatives in supply chains. As post-secondary students increasingly own laptop computers, the availability of substantial personal computing power allows serious games based on discrete-event simulation to be integrated into educational programs. This paper describes such a game, called *The Supply Chain Game*, how that game was constructed, how it is used in a post-secondary supply chain course, and the outcomes of using the game over a six-year period. Immersing students in a virtual supply chain has proven to impart months or years of practical experience in a very short time. Moreover, the real context of the game motivates students to learn course topics taught using more traditional pedagogy by highlighting the relevance and importance of those topics in managing real businesses. This account offers a strategy for educators to develop similar games.

1.0 INTRODUCTION

Serious gaming has gained a foothold in many areas of education, including military education (Waldman 2009), medicine (Verdaasdonk 2009), health education (Skiba 2008), training for surveillance of potential natural disasters (Harteveld; Guimarães, Mayer, & Bidarra 2010), encouraging an affinity toward humanitarian aid efforts (Oliver 2010 and Joseph 2010), in stroke rehabilitation (Burke, McNeill, Charles; Morrow, Crosbie; & McDonough 2009), flight simulation, and in many other realms. Another natural context for serious games is in business education and, in particular, in supply chain management. This is so, for one reason, because of the ample simulation methodologies that have been developed over the past fifty years in operations management. These methods can be deployed with a “re-packaging” of the user interface that guides the user’s (student’s) interaction with the simulation-based game. Moreover, introducing serious games in the realm of supply chain management is relatively easy because mimicking the typical supply chain decision-making environment is easy because business decision support and transactional systems use visual display devices, which are inexpensive and readily available. In comparison, the effective implementation of serious games in other contexts requires

development of cutting edge and potentially expensive user-interface technologies such as haptic devices and caves to replicate accurately the target environment.

This paper describes one such serious game, its purpose, and how it was specifically designed to support the educational goals of an existing supply chain management course. That game is called “The Supply Chain Game,” and was developed by the author for use in courses at the Mason School of Business at the College of William and Mary. It is an Excel-based simulation game where the students make day-to-day decisions about how to manage a supply chain in competition with other students.

2.0 BODY

Monte Carlo simulation and discrete-event simulation are the most commonly applied simulation methodologies in business contexts where events are uncertain. When supply chain decisions are analyzed, discrete-event simulation is the most appropriate of these methods because we must necessarily be concerned with the flow of goods in the system over time and how the state of the supply chain evolves over time, contrary to Monte Carlo simulation where the state of the system is “wiped clean” before generating another

“snapshot.” We are concerned with the development of a serious game in supply chain management using simulation as a basis, and so we focus on discrete event simulation.

The primary technological concern in constructing a discrete-event simulation is the appropriate modeling of uncertain events, which, in the context of supply chains includes (i) the quantity of goods produced, (ii) the quality of the goods, (iii) the reliability of machine and human resources used to produce goods, (iv) the time required to transport goods, and (v) the quantity of goods that will be demanded by customers. Fortunately, the theory of creating pseudo-random variables is well developed and increasing computation speeds allow many such random variables to be generated in a very short time. Moreover, simulations run as an educational exercise do not necessarily need to be large scale simulations such that computational limits are challenged. To the contrary, simulations for educational purposes can be simple so that the more technical aspects of simulation need not be considered, such as complex variance reduction techniques. This is certainly true for simulation that we present here, which supplements teaching at the introductory level of supply chain management.

2.1 Motivation for the Supply Chain Game

Prior to developing The Supply Chain Game, the author had been teaching supply chain management courses using common pedagogical techniques such as lectures, case studies, assignments, and tests. Lectures for supply chain techniques were, as is typical and appropriate, mathematically based. The core topics of the author’s course dealt with control of the flow of goods through supply chains, and on these topics specifically:

- Forecasting
- Inventory ordering and replenishment decision models

- Economics of transportation (e.g., trucking and rail costs)
- Shop floor scheduling
- Aggregate planning

The course received high evaluations, but the instructor feared that the students’ grasps of the topics might be limited by their comfort level with math, and that they might not be seeing the linkage between all the methods above and how, together, they are used in supply chain management. The author hypothesized that a computer-based game that simulated a supply chain could show how all the supply chain decisions are interrelated, demonstrate the importance of competent supply chain management to a company’s financial success, and provide a context for all the course topics. In addition, if the game was fun, it might also engage the students such that they would be more motivated to study the mathematics inherent in supply chain management decision analysis.

2.2 Game Design

The Supply Chain Game was designed so that understanding all the aforementioned course topics was necessary to score well on the game. Accordingly, the game’s supply chain includes these supply chain links through which one stock keeping unit (SKU) flows:

- A manufacturer
- Transportation from the manufacturer to a distribution center
- A distribution center

Goods are ordered by the distribution center from the manufacturer in either regular or expedited modes. The economics of the regular shipping mode emulate the economics of trucking: a fixed cost is incurred for each shipment, and a variable cost is added for each additional unit shipped on the truck. The truck has a limited capacity. Expedited shipments are assessed only a variable cost for each unit which is much higher than the variable cost for trucking to reflect express delivery

options such as FedEx, UPS, DHL, or other carriers. Regular shipping takes six days whereas expedited shipping takes one day. These are transport times, and they do not include the time required to manufacture the goods in the order.

Players make an ordering decision each day, which is transmitted to the manufacturer. The manufacturer processes these orders in the sequence they are received, except that expedited orders are given priority over regular orders. A particular order is not shipped until all the units in that order are complete. The manufacturer has a finite capacity, which limits the number of units per day it can produce.

Customers are served from the distribution center, which receives the goods from the manufacturer. Each day an uncertain quantity of units is demanded, which are determined by standard random variable generation techniques. (Common random variables are used for all students in a class so that the competition among students is fair.) The game is played over a two-year period, or 730 days. The mean demand is not stationary, but varies with the selling season. This feature was intended to motivate students to apply forecasting methods for seasonal demand; in this course Winter's model is taught, which is appropriate for the game data.

Transportation from the manufacturer to the distribution center takes 6 days for regular orders and 1 day for expedited orders.

Students must also make a decision about expanding the capacity at the manufacturer given the historic demand data. Increasing the manufacturer's capability to make an additional unit each day requires a specified investment. This decision motivates the students to use the aggregate planning model that is taught in the course. The transportation economics lectures in the course and well as the inventory decision analysis methods help students understand what the best ordering policies are.

The goal of the game is to execute these decisions in a manner that minimizes cost. The other key performance measure that is tracked in the game is the percentage of orders that are fulfilled on time at the distribution center (i.e., "service level").

2.3 Game Platform

The Supply Chain Game is written in Excel with Visual Basic for Applications (VBA). The motivation for this choice is threefold. First, the students are very familiar with Excel, so this is a comfortable platform. Second, future courses in the supply chain curriculum involve training in VBA, so this serves as an introduction of how Excel can be used for simulation and other more general decision models. Third, coding the game was a relatively small time investment for the instructor.

2.4 Features of the Game

The image of the supply chain is shown in Figure 1. Students can see the order and processing status at the manufacturer, the status of the two delivery pipelines (regular and expedited modes), and the inventory status at the distribution center. The students are given two years of demand history before they play the game the first time. As they play the game a graph updates the demand history on an Excel worksheet to reflect realized demand, as shown in Figure 2. This worksheet also shows key performance measures as the game progresses. The students make inventory policy decisions by entering them into the supply chain worksheet (Figure 1), and then executing those policies for a selected number of days using the dialog box shown in Figure 3. (Multiple days can be executed using the same policy, if desired.) That dialog box also facilitates capacity decisions. Figure 4 shows a worksheet within the game spreadsheet that describes game instructions and also the costs of capacity, shipping, production, holding cost, and the financial impact of missing a delivery to the customers at the distribution center.

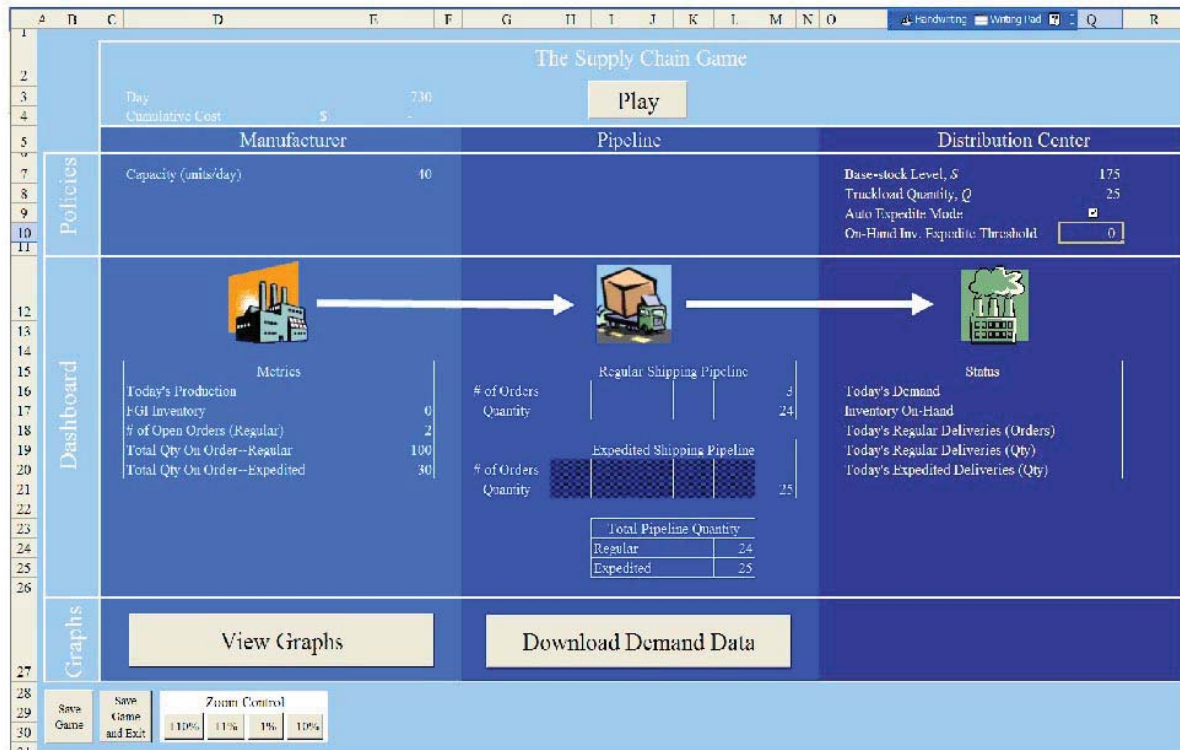


Figure 1: The Supply Chain Game Supply Chain

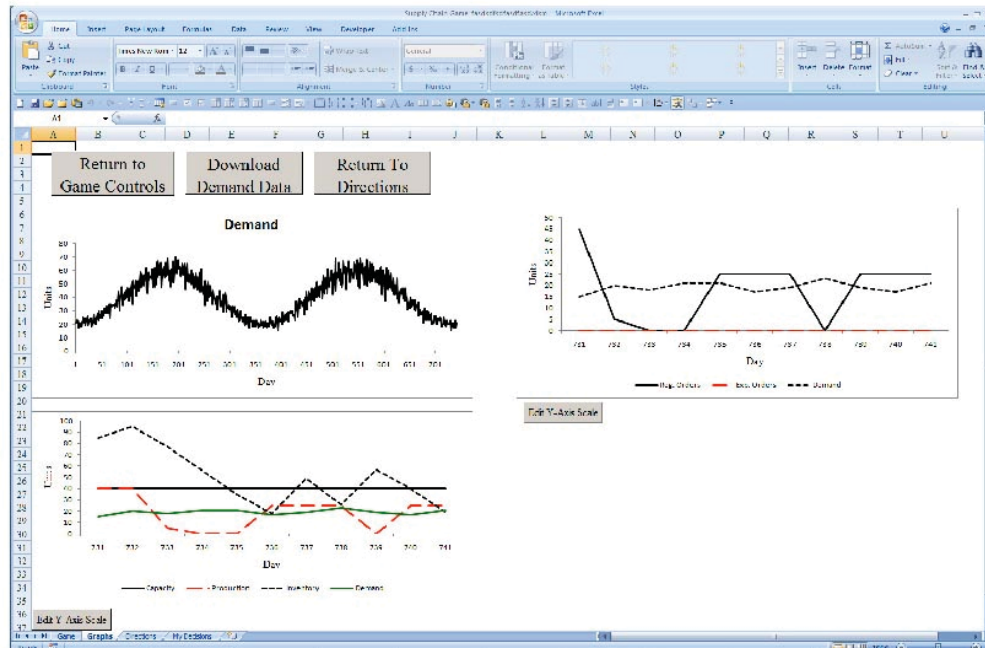


Figure 2: Demand History and Performance Metrics

The Supply Chain Game User Screen

The Supply Chain Game

Day

of Days to Operate

Expand manufacturing capacity? ☒

Enter the capacity expansion increment:

This expansion will take effect on day:

Figure 3: Decision Dialog Window

The Supply Chain Game

Setting

The setting for the Supply Chain Game is a three-echelon supply chain consisting of a manufacturer, a distribution center, and end customers. The manufacturer produces goods, which it ships to the distribution center via regular and expedited shipping modes according to the policies set at the distribution center. The manufacturer has a capacity limit on how many units of goods it can produce per day. The game unfolds on a day-by-day basis. The game reveals, each day, how much demand is realized, what orders are placed by the DC, how many goods the manufacturer produces, and how many goods are shipped. Demand is exogenous (it is out of your control as a manager), while the other events occur as a result of the management policies at the DC and the manufacturer.

Goal

Your mission is to set the policies by which the supply chain is managed in order to **minimize** the total cost of operating the supply chain over a **two-year period** (730 days). Supply chain costs are the sum of costs to manufacture, move, store, and, ultimately, deliver the goods to the end customer. This cumulative supply chain cost accrues according to the various cost components listed below.

Cost Item	Cost	Units of Measure	Notation Used in Class	Cost Category
Backorder Cost	100.00	\$ per unit per day	b	Inventory
Holding Cost	0.75	\$ per unit per day	h	Inventory
Truckload Cost	25.00	\$ per truckload	c_{TL}	Shipping-Regular Mode
Unit Shipping Cost	0.05	\$ per unit	c_u or $c_{u,TL}$	Shipping-Regular Mode
Unit Shipping Cost	20.00	\$ per unit	c_e	Shipping-Expedited Mode
Unit Manufacturing Cost	30.00	\$ per unit	c_p	Manufacturing
Unit Manufacturing Capacity Cost	20,000.00	\$ per unit of capacity	c_{pc}	Manufacturing

Your Role as Supply-Chain Manager

A player's performance is dependent on their decisions, which include...

Base stock policy, S

When the *inventory position* falls below the base-stock level, then an order is issued to bring the inventory position back up to the base-stock level. However, the "issued" orders are held until the cumulative number of orders equals or exceeds the truckload quantity.

(The DC's *inventory position* is the quantity of goods physically on hand, which could be negative if goods are backordered, plus the quantity of goods in the orders that have been placed but not yet received.)

Truckload quantity, Q

This is the number of units that you want to put in each truckload that is sent from the Manufacturer to the DC. A maximum of 100 units will fit on a truck.

Expedited orders

Orders can be sent from the Manufacturer to the DC using a faster mode of transportation. Expedited delivery differs from the regular delivery mode also in these dimensions:

- Orders for less than a truckload quantity can be ordered via expedited delivery. The fee structure is on a per-unit basis rather than a truckload basis.

Figure 4: Game Instructions and Cost Parameters

2.5 Integrating the Game and the Course

After the first session of the course, which introduces the general concept of supply chain management and the course's logistics, the students play The Supply Chain Game out of class as a "pre-test" and the results of the game are debriefed in the next session. At this point in the course, students have been taught no supply chain skills, so they are forced to play the game intuitively. The debrief discussion points out all the decisions that are important in the game and invariably the students voice a discomfort with not having good methods for determining better policies than dictated by their intuitions. The instructor, then, mentions that the course will cover models and analysis to help make better decisions. Thereafter, each course session that addresses one of those models can cite The Supply Chain Game as motivation for the importance of each topic, and as a tangible context for the models.

The last assignment in the course is to take the demand data from the original play of The Supply Chain Game, and analyze it using the tools taught in the course to determine the decision policies that should be used going forward (i.e., ordering policies, transportation policies, and capacity expansion policies). Students present these analyses in class and the instructor posts each student's analysis so that the students may use each other's analysis.

In the final class assignment, students, then, play the game a second time over a simulated two-year horizon using their analyses. This experience is followed up with a de-brief session. During the course, and before this second opportunity to play the game, students usually realize that the original game is not structured so that the optimal policy is easily implemented. Thus, the instructor provides a way to do this in the second round of the game, as well as providing an "autoplay" mode, where the students can play the entire two years by

specifying their non-stationary decision policies on a second worksheet and executing the two-year sequence by clicking the "Go" button only once. In this case, the game implements the students' policies automatically each day rather than a student having to input a decision day by day.

3.0 DISCUSSION

3.1 Game Results

On average, students cost performance in The Supply Chain Game decreases 25% from the first to the second play, which is a non-trivial improvement in real-world contexts. Typically, the greatest cost reduction in a class of about 20 students is 90%.

The range of supply chain cost performance among the students in the class is typically significant so that it is apparent that many of the students' supply chains would be out of business based on traditional profit margins. This observation emphasizes the importance of effective supply chain management.

Past students have made these comments about the game:

"As a non-numbers oriented student, Professor Bradley's self-created supply chain game was the only tool that allowed me to grasp the moving parts of supply chains. His game was simple to use but incredibly educational and fun. Because of his game, I can intelligently discuss supply chains and things that impact them."

"The Supply Chain Game was an excellent learning tool. I was able to apply and test the principles that we learned in the lectures."

"I want to thank you for a great class this semester. The class definitely exceeded all my expectations. The supply chain game simulation really helped me understand fundamental concepts with

all the necessary details behind supply chain decisions. I am just waiting for my next opportunity at work to use the concepts that I have learnt from you.”

“I found Professor Bradley’s use of The Supply Chain Game as a tool for analyzing and testing the techniques we learned during the Supply Chain Management course to be an extremely powerful (and dare I say, entertaining) way to drive home how useful the spreadsheet modeling techniques learned in class can be in helping the supply chain manager with creating policies to benefit his or her company’s bottom line.”

“The best part of [Bradley’s course] was an Excel-based supply chain game. We played the game initially during the first week of the course when we had not learnt any key concepts of supply chain. The first time when we played the game, our numbers were very bad. From then on, when Prof. Bradley taught supply chain concepts he would relate them to the game. That approach helped us to have a better understanding of what we did wrong and how we can make things more efficient. We then played the game at the end of the course, our numbers improved significantly.”

The author, anecdotally, has also perceived an increased interest in the course sessions and, once students finish the course, the students often reference The Supply Chain Game, possibly indicating that the game and the course have made long-lasting impressions. Of course, these observations have not been collected or analyzed scientifically and, even if accurate, the causes of these anecdotal observation could be rooted elsewhere, such in a changing student body demographic that has more interest in supply chains or more analytical capability.

3.2 Lessons Learned

The students learn how to apply the supply chain tools taught in the course through playing The Supply Chain Game, but the following are some of the more specific lessons that the students learn through the game. These are lessons that would be difficult for students to comprehend if addressed only in mathematical terms. The Supply Chain Game, in contrast, makes these lessons very real for the students.

The mathematical models that determine how many items to order each day depend on the time between when an order is placed to when it is received at the distribution center (i.e., the “lead time”) and the statistical variation of lead time. It is intuitive for people to think only of transportation time as being the lead time. Many other factors can increase the lead time, however, including the time to transmit the order to the manufacturer, delay at the manufacturer, and time to unload a truck. The supply chain game includes the potential for additional delay at the manufacturer when it is backed up with orders. Playing the game leads to this realization and students realize the necessity for including this time component in their calculations (as well as its variation). The students also come to realize that their decision about how much to increase the manufacturer’s capacity affects the manufacturer’s backlog of orders, and hence also the lead time, the lead time variation, and, hence, the ordering decision. This is a big lesson on how all the supply chain decisions (in this case capacity and inventory policy) are linked.

The expedited delivery option is offered to the students in the game “as a trap.” The author has witnessed many instances in industry where upon running out of inventory managers expedite deliveries from suppliers at great expense due to the importance of fulfilling customer demand at the distribution center a high percentage of the time (i.e., service level). This game demonstrates that while expense increases

with expedited deliveries, on-time delivery performance does not improve significantly, and many times service level performance actually decreases because the expedited orders delay the orders in the pipeline that are to be delivered in regular delivery mode. Inventory models exist that consider both regular and expedited deliveries, but teaching the math involved with these models requires relatively sophisticated calculus, which is beyond the level of a majority of MBA students. The Supply Chain Game, thus, makes this lesson observable for this target audience.

The Supply Chain Game also demonstrates to the students the tradeoff between capacity and inventory. The students realize that they can satisfy demand at the distribution center on time by either having larger inventories or a larger capacity at the manufacturer. This alternative calls into question which of these alternatives most economically provides a high service level. The aggregate planning model covered in class provides a framework to evaluate the alternative combinations of inventory and capacity.

4.0 CONCLUSION(S)

Indications are that The Supply Chain Game has increased the efficacy of the author's supply chain course, resulting in greater interest in the topic of supply chain management and greater motivation to learn the mathematic decision techniques that are important for companies to remain competitive in the marketplace. This game allows students to glean months, if not years of experience in a very short time and, therefore accelerates students' learning when they take a permanent job or an internship because the game gives them an impression about how all the pieces of the supply chain game fit together and how supply chain decisions are best made.

5.0 REFERENCES

Burke, J.; McNeill, M.; Charles, D.; Morrow, P.; Crosbie, J.; McDonough, S. "Optimising Engagement for Stroke Rehabilitation using

Serious Games," *Visual Computer*, December 2009, Volume 25, Issue 12, 1085-1099.

Harteveld, Casper; Guimarães, Rui; Mayer, Igor S.; Bidarra, Rafael. "Balancing Play, Meaning and Reality: The Design Philosophy of LEVEE PATROLLER," *Simulation & Gaming*, June 2010, Volume 41, Issue 3, 316-340.

Joseph, Barry. "Get Serious: Games that Address Contemporary Social and Global Issues," *School Library Journal*, February 2010, Volume 56, Issue 2, 24-25.

Oliver, Sophie. "Simulating the Ethical Community: Interactive Game Media and Engaging Human Rights Claims," *Culture, Theory & Critique*, April 2010, Volume 51, Issue 1, 93-108.

Skiba, Diane J. "Games for Health" *Nursing Education Perspectives*, Jul/Aug 2008, Vol. 29, Issue 4, 230-232.

Verdaasdonk, Egg; Dankelman, J.; Schijven, M. P.; Lange, J. F.; Wentin, M.; Stassen, L. P. S. "Serious Gaming and Voluntary Laparoscopic Skills Training: A Multicenter Study." *Minimally Invasive Therapy & Allied Technologies*, August 2009, Vol. 18, Issue 4, 232-238.

Waldman, Andrew. "Serious Games," *National Guard*, November 2009, Vol. 63, Issue 11, 26-29.

6.0 ACKNOWLEDGMENT(S)

The author thanks the many students who have responded to the author's requests for feedback and comments on The Supply Chain Game, and who have made teaching supply chain management such a fulfilling experience.

4.4 Effects of a Pedagogical Agent's Emotional Expressiveness on Learner Perceptions

Effects of a Pedagogical Agent's Emotional Expressiveness on Learner Perceptions

Enilda J. Romero
Old Dominion University
eromero@odu.edu

Ginger S. Watson, Ph.D.
Old Dominion University
gswatson@odu.edu

Abstract. The use of animated pedagogical agents or avatars in instruction has lagged behind their use in entertainment. This is due in part to the cost and complexity of development and implementation of agents in educational settings, but also results from a lack of research to understand how emotions from animated agents influence instructional effectiveness. The phenomenological study presented here assesses the perceptions of eight learners interacting with low and high intensity emotionally expressive pedagogical agents in a computer-mediated environment. Research methods include maximum variation and snowball sampling with random assignment to treatment. The resulting themes incorporate perceptions of importance, agent humanness, enjoyment, implementation barriers, and suggested improvements. Design recommendations and implications for future research are presented.

1.0 INTRODUCTION

Increased access to the Internet and widespread use of computers in the last decade has transformed computer-mediated instruction. Today there are a variety of different formats to support this type of educational setting such as web-based instruction, computer-based instruction, computer-based instructional simulations and serious games. In all of these formats the instructional content is presented in a combination of text, images, animations and/or audio representations. However, in an effort to increase human-computer interaction and to create the illusion of human to human interaction, these instructional environments have also seen an increase in the use of animated pedagogical agents.

Animated pedagogical agents are life-like computerized characters that facilitate learning in interactive environments [1, 2]. Some pedagogical agents are used as support tools like virtual tutors or help aids. Others are used to present the instructional content like a virtual instructor. Additionally, some animated pedagogical agents perform as virtual actors. An example of such function is CATHIE (Computer Agent

Teaching Helping Interactions Effectively), an animated pedagogical agent portraying the role of a student seeking advice from a human services employee [3].

In computer-mediated instructional environments, pedagogical agents can be represented as anthropomorphic characters (i.e., non-human like cartoon images) or human-like images or 3D figures. Overall, animated pedagogical agents should meet the basic interface agent characteristic of agency, responsiveness, competence, and accessibility as well as, the character building qualities of animations, gestures, communication, and personality [4]. Of course, in order for instructional designers to recreate the essence of life by developing characters that seem to think, live, and feel it is important to capture the essence of humanity by providing appropriately timed and clearly expressed emotions [4].

The focus of this investigation is to expand current knowledge on the role of animated pedagogical agents in computer-mediated instructional environments. Of particular concern are the effects of emotion (low versus high) intensity on learners' perceptions towards human-like emotionally expressive animated pedagogical agents.

2.0 LITERATURE REVIEW

Prior research on animated pedagogical agents has focused mainly on quantitative inquiries questioning the animated agents' effect on learning and the learners' perceptions of animated agents reported through survey instruments [3, 5, 6, 7, 1, 8, 9]. Some of these studies obtained positive results for both learning and learners' perception of the animated agent [9, 5]. However, other studies failed to obtain significant results for either variable [3, 1]. In the majority, the studies reported mix results in which learners' gained a positive perception of the animated agent but performed poorly on the learning tasks [6, 7, 8].

In the case of qualitative inquiries, research on animated pedagogical agents has been limited to a single phenomenological exploration that focused on the experiences individuals have when holding a dialogue with conversational pedagogical agent for educational purposes [10]. In this phenomenology, analysis from interviews and written reflections yielded five main themes (i.e., asking questions and trying to understand the pedagogical agent, the sense of time, body reactions, sense of place and humanizing the pedagogical agent). Overall, the findings support that individuals interact with virtual characters as if they were interacting with other humans [10]. Participants of this phenomenological exploration reported losing track of time, testing the knowledge of the animated agent, paying close attention to both utility and usability but also the interaction design related to aesthetics and emotion.

So far, only two studies utilized human-like emotionally expressive animated agents while measuring the agent's effect on learning and learners' perception [11, 12]. Kim et al. [11] used a quantitative approach to investigate learners' perception and learning outcomes using verbal and facial expression of emotion and gender in the interaction between an animated pedagogical agent and the learners.

Overall, results indicated that only the positive emotional expression (male and female) increased learners' social judgment of the agent and the interest of the learner. No significant learning outcomes were obtained in the study from any of the treatments.

Veletsianos [12] used a mixed method approach in a study in which learners interacted with either a verbally expressive or a non-verbally expressive animated pedagogical agent. The quantitative results of the study reported significant learning outcomes and positive learners' perception of the animated agent. Additionally, the qualitative data collected served as a method to gain a deeper understanding of the learners' experiences as they interacted with the animated agent. In total six main themes emerged from the data: expressiveness, multidimensional and contrasting agent perceptions, items of distraction, pedagogical agent affordances, use of 'human' as a measure of comparison, and agent focused with the lesson on the periphery. Overall, the comparison of qualitative data from both treatments revealed that change in verbal expressiveness qualities influences student perceptions and learning outcomes and that agent expressiveness is an important element of pedagogical agent design [12].

Due to the limitations of the past qualitative research on the topic of animated pedagogical agents, which only focused on dialogue experiences with a conversational agent and the verbal emotional expression of an agent, it is important to further explore in a qualitative manner the topic of animated pedagogical agents. Perhaps, the qualitative research on animated pedagogical agents can help understand the mixed results (positives and negatives) obtained by past quantitative inquiries. Or, it could also provide critical design feedback as experienced by the learners in their interactions with the animated pedagogical agents. Nonetheless, the qualitative research needs to center its attention on

animated pedagogical agents with both verbal and facial expressions of emotion.

3.0 PURPOSE STATEMENT AND RESEARCH QUESTIONS

The purpose of this qualitative study was to describe the lived experiences of the learners as they interact with an emotionally expressive animated pedagogical agent in a computer-mediated instructional environment. Specifically, the researcher focused on the lived experiences between learners who interact with a low versus a high intensity emotionally expressive animated pedagogical agent. The main research questions that guided this study were: how do learners perceive low intensity emotionally expressive agents? (Research Question 1) And, how do learners perceive high intensity emotionally expressive agents? (Research Question 2)

4.0 OPERATIONALIZED VARIABLES

Low and high intensity emotion variables were operationalized based on the programming of the facial and verbal expressions of the animated agent.

4.1 Low Intensity Emotion

The animated pedagogical agent had a soft emotional tone in the voice. Eye brown, mouth, and cheek movements were subtle and gentle. Head movement was limited and delicate.

4.2 High Intensity Emotion

The animated pedagogical agent had a strong emotional tone in the voice, with very high pitch and intonation. Eye brown, mouth, and cheek movements were exaggerated and excitement-like. The animated agent experienced a significant amount of head movement.

5.0 RESEARCH TRADITION

For this research study, a phenomenology tradition was utilized. Phenomenology is a qualitative research paradigm focused on

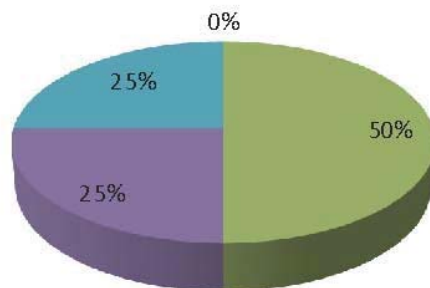
capturing the essence of the phenomenon under study. It descriptively captures the invariant characteristics of the phenomenon; and clarifies its significance and organization [13]. Under phenomenology, researchers participate in intentional analysis, in which the phenomenon under study is experienced by the participants as it normally occurs and it is reflectively explored to truly understand the experiential process through which the phenomenon is lived [13]. In addition to exploring the essence and intentional analysis of an experience, phenomenology encourages researchers to abstain from integrating theories, explanations, and hypotheses about the phenomenon under investigation [13]. Instead, the researcher learns about the phenomenon subjectively without any naive or irrelevant beliefs.

This qualitative inquiry captured the essence of the experience using the learners' perspectives of the agent and this type of instruction. The goal was to recognize commonalities and differences among the learners' experiences between the two treatments (i.e., low intensity and high intensity emotion).

6.0 METHODOLOGY

6.1 Participants

The participants of the study were eight university students, both undergraduate (i.e., six students) and graduate level (i.e., two students), from a public university in southeastern United States. Five of the participants were females and three were male students. Six participants were within the 18 to 25 age group and two were in the 25 to 35 age group. Most of the participants were full time students; only one participant was a part time student.



According to the responses in the demographic questionnaires, all participants were either competent, proficient, or expert users of computers. Also, they were very familiar with desktop and laptop computers and knew how to use Windows and Mac operating systems. On average, the participants spend 18 hours using the computers at home and 17 hours of computer usage at works in a weekly basis. Typically, the participants used the computer for word processing (15%), email (17%), reading the news (13%), research (11%), and social networking (15%).

6.2 Sampling Strategies

For this phenomenological inquiry the researcher utilized a purposive sampling method, this form of sampling method provided information rich cases that yield insightful and in – depth understanding of the phenomenon under study [14]. Of the different strategies that can be used for purposively selecting information rich cases two specific sampling strategies were used: maximum variation sampling and snowball sampling. The aim of using maximum variation sampling is to capture and describe the central themes that cut across a great deal of variation [14]. In this manner the common patterns that emerged from

both undergraduate and graduate student participants are of particular interest and value in capturing the core experience and central, shared dimensions of emotionally expressive animated pedagogical agents in computer – mediated instruction.

In the case of snowball sampling, the process begins by asking well-situated people who the researcher should speak with [14]. In this phenomenology, snowball sampling was used to locate information–rich participants. This was done by asking the initial participants in the study to pass on information of the study to other undergraduate and graduate classmates.

In order to obtain participants that fit the characteristics of both purposive sampling strategies, recruitment notices were send out through university and departmental listserv, university social networking sites, and university bulletin board fliers. No specific age, gender, sexual orientation or ethnic/racial group restriction were set for selection of the participants. As compensation, participants received a five dollar gift card to a local coffee shop.

6.3 Materials

6.3.1 Instructional Environment

The instructional environment consisted of a linearly sequenced set of web pages that provided an overview on the topic of African history. Participants were first presented with an introduction and learning objectives for the lesson. Then, they progressed through thirteen pages of instructional content. After the instructional content, learners advanced to the conclusion. In order to move forward and backwards, participants used the “Previous” and “Next” buttons on the top left and right corners of the web pages. Participants used the progression menu, located on the left side of the web page, to track how far they had advanced through the lesson content.

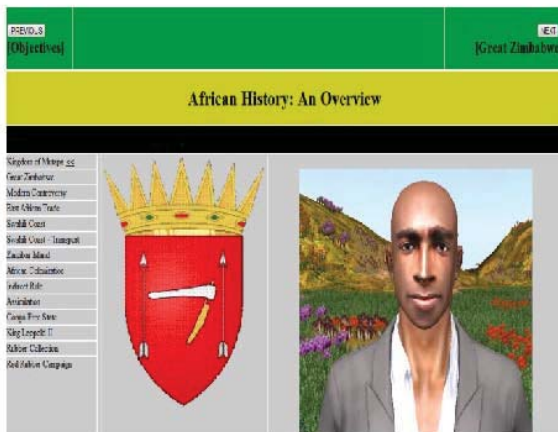


Figure 1. Instructional Environment



Figure 2. Animated Pedagogical Agent

In each instructional content page, the participant would see the title of the lesson (i.e., “African History: An Overview”), the progression menu, the title for the content of the page (e.g., “Kingdom of Mutapa,” “Great Zimbabwe,” “Zanzibar Island,” etc.), an image related to the content of the page and the animated pedagogical agent.

The instructional environment was developed using Adobe Dreamweaver. The instructional content was prepared by a subject matter expert on the topic of African history.

6.3.2 Animated Pedagogical Agent

The animated agent presented in the instructional environment consisted of a young (25 – 30 years old) male with African American physical attributes. The agent wore a white shirt with a gray sports jacket. It used semi-formal language and spoke with a voice relatable to the age range. The background of the animated agent video presented of a hillside green area with bushes and flowers.

6.4 Data Sources

Data was collected through individual interviews, guided reflection journals, and a demographic questionnaire.

6.4.1 Individual Interviews

The interviews lasted 20 – 35 minutes each and occurred immediately after the participants interacted with the emotionally expressive animated pedagogical agent. They had a guided format and consisted of open ended questions. Participants were asked to describe their experiences as they interacted with the computer-mediated environment. The aim was to gain a deep understanding of the participant’s perceptions of the emotionally expressive animated agent. In order to have complete accuracy during the transcription process, the interviews were video recorded.

6.4.2 Guided Reflection Journals

Similar to individual interviews, guided reflection journals were completed after the participant interacted with the animated agent in the computer-mediated instructional environment. The journals were written electronically. Specifically, the researcher provided an electronic form with comment/essay boxes in which participants typed their reflections on the experience. The reflections of each participant were

guided by a series of questions. Participants had 40 – 50 minutes to type their guided reflection journals.

6.4.3 Demographic Questionnaire

Lastly, participants were asked to complete a demographic questionnaire. The questionnaire was provided in paper format and consisted of a single sided page with eight questions. Some of the data collected through the demographic questionnaire were: gender, age group, college education level, type of student (full-time or part -time), computer experience, computer usage (amount of time), and type of computer usage (e.g., social networking, gaming, research)

6.5 Procedures

After initial contact with the participants, the researcher sent an email to the participant with a link to register for the study and select a specific time slot for participation. The day of the session, the participants were asked to sit comfortably in front of a computer monitor with a wireless keyboard and mouse. The researcher proceeded to brief the participant on the purpose of the study and the structure of their participation. Various consent forms (i.e., participation and video consent forms) were provided to the participant. The researcher gave the participant time to read over the consent forms and sign if they felt comfortable with their participation in the study.

Once the participants had signed the consent form, the researcher randomly assigned the participant to a computer-mediated environment (i.e., low intensity emotionally animated agent or high intensity emotionally expressive animated agent). At this point, the researcher left the room and asked the participants to open the door once they had completed the interaction. After the researcher re-entered the research room, the participant was randomly selected to complete either a guided reflection journal or an interview protocol.

Before leaving the research room, the participants completed the demographic questionnaire and the gift card receipt form. In total, four participants were randomly assigned to the high intensity emotionally expressive animated agent and four to the low emotionally expressive animated agent.

6.6 Data Analysis

6.6.1 Individual Interviews

As each interview occurred, they were transcribed and analyzed. Video and voice recording of the interviews was used to transcribe the dialogue between the participant and the researcher. The transcriptions consisted of exact verbatim.

In addition to the interview transcripts, a contact sheet was filled after each interview. The sheet was filled by the researcher and it included four main questions. The questions were used to make the researcher reflect on the interviewee, the interviewee's responses, important issues or themes that emerged during the interview, and to compare the interview with other interviews.

6.6.2 Guided Reflection Journals

To analyze the guided reflective journals, each participant's journal copied from the electronic format to a MS Word format. Exact verbatim was copied. Then, for each journal a contact sheet was completed.

The contact sheet used to analyze the guided reflection journal was the exact same format and questions as the contact sheet used for the individual interviews.

6.6.3 Demographic Questionnaire

To analyze the demographic questionnaires, an MS Excel file was created. Responses to each question were coded in the columns/rows of the spreadsheet (i.e., gender, age group, education level, type of student, computer experience, computer usage, and type of computer usage). Totals were calculated

for each response option and graphic pie charts were created for each question category.

After all the interviews, guided reflection journals and demographic questionnaires were analyzed, two members of the research team conducted a phenomenological analysis of the data. The intent was to grasp the meaning, structure and essence of the lived experiences of the participants [14] as they interacted with the emotionally expressive animated pedagogical agent in the computer – mediated environment. To start the process, the researchers Epoche. This means that the researchers bracketed or restrained from judgment, personal bias, or eliminated any involvement with the subject material [14].

The next step was the phenomenological reduction of the data. To reduce the data, all interviews and reflection journals were coded separately by two researchers. During the coding process, key words and phrases were pulled from the data to develop initial codes. Code lists from both of the researchers were then compared against each other for any similarities or discrepancies. All the codes were entered into a codebook. The final version of the codebook included 19 keywords and phrases that were grouped into five coding categories.

Following the reduction of the data, the horizontalization process occurred. The purpose of horizontalization was to spread out the data and to organize it into meaningful clusters, in which irrelevant, repetitive or overlapping data are eliminated [14]. Five main themes were identified from the horizontalization process. A MS Excel spreadsheet was created to enter textual and structural descriptions for each of the five main themes.

7.0 RESULTS

Five themes emerged from the individual interviews and guided reflections journal. It

is important to mention that these themes are mutually supporting and overlapping. They will be described in turn.

7.1 Perceived Importance

Participants in both treatments, low and high intensity emotion, described reasons why emotionally expressive animated pedagogical agents are important in computer – mediated environments. Participant B, assigned to the low intensity emotion treatment, mentioned that “it is kind of nice to have a face because you can pick up some cues if something is a little bit more important, that needs to be paid close attention to.” Similarly, a participant in the high intensity treatment provided an example of how emotionally expressive animated agents could be implemented to support the learning experience of the students. Participant E felt that “emotions are a huge part of people and that if you take that out of a teacher it would be harder for the students to relate and they will get bored. There is just not enough stimulation for the student [without emotion].”

However, one participant mentioned that they could not see the difference between using an emotionally expressive agent and just videotaping a real person. This participant did not consider emotionally expressive animated agents necessary or important to computer-mediated instruction.

7.2 Perceived Enjoyment

Regardless of agent expressiveness, low or high intensity, participants were both critical and complimentary regarding their perceived enjoyment of the animated agent.

Some of the complimentary comments from the participants in low intensity treatment were related the facial expressions of the agent. Participant D mentioned that “he likes that sometimes the avatar would stop and smile. Also, raising the eyebrows every once in a while was a nice touch.” Similar complimentary comments were reported by the participants in the high intensity emotion

treatment. Participant G mentioned that he liked the sharpness of the agent's appearance. And, just like participant D, he enjoyed the fact that while speaking the avatar's facial expressions and head movements were interacting simultaneously with the talk.

Some of the participants in the high intensity treatment did not enjoy interacting with the agent. Participant E mentioned that "the agent didn't have much to look at" and, similarly, participant F stated that "she found himself looking at the pictures more than looking at the avatar." Participant F also mentioned that although she found the topic of lecture interesting she did not enjoy looking at the agent emotional expression.

7.3 Perceived Humanness of the Agent

All participants (both treatments) perceived the agent as "professional looking", "sharp dresser," "very modern" and someone that "looked like a professor."

However, participants in both treatments also revealed that the humanness of the agent was easily discredited by lack of proper facial expression of emotion or lack of synchronization between audio and mouth movements. Participant C, a participant in the low intensity emotion treatment, mentioned in her interview that "although he was talking about an upsetting subject, I did not recall any sad facial expression from him. I saw that he was smiling briefly." In the high intensity emotion treatment participant F mentioned there was a lack of synchronization, "I noticed that sometimes the voice and the movement of the mouth did not quite line up. For a second it discredited the avatar."

7.4 Perceived Barriers for Implementation

For the participants in the low intensity emotion treatment, as well as one participant from the high intensity emotion treatment, the main barrier for the

implementation of emotionally expressive animated agents was the fact that they are too distracting. One participant stated, "I had to focus on what it (the agent) was saying because his face distracted me from his voice." Participant D, who was assigned to the low intensity emotion treatment, also mentioned how distracting the agent was for him. The participant said, "The funny faces he made with the smile and moving eyebrows distracted me into watching them instead of listening to what he had to say." Participant C too agreed with the level of distraction related to the avatar, he mentioned "I focused more on what his face was doing than what was coming through the speakers." Participant E, who interacted with the high intensity emotion agent, mentioned "this could be extremely distracting to students because they would spend more time focusing on why the avatar's expression was a certain way rather than the lesson." Participant G felt that "because avatars are not actually people, their emotions aren't displayed as smoothly as ours are; therefore, glitching and delays can lead to student distraction, which could ultimately cause a barrier in learning."

7.5 Suggestions for Improvement

Data from the interview transcripts and guided reflections journals indicated that some participants would be glad to interact with an emotionally expressive animated pedagogical agent if given the opportunity. However, they would like to see some improvements and changes to the animated agent. For example, participant D who was also assigned to the low intensity emotion treatment suggested additional work on the facial expression component of the animated agent. He stated, "I could easily spot the same pattern of movements and expression from the avatar. A richer library of expressions would make a great difference."

One participant assigned to the high intensity emotion treatment recommended

having more emotional inflection in the voice of the animated pedagogical agent. Participant G said, "Sometimes I imagined a person reading a script."

8.0 DISCUSSION

Across all themes, the results indicated that participants in both treatments have similar perceptions of animated pedagogical agents. In fact, there are more similarities in perception than there are differences. Overall, the learners in the low emotion intensity treatment perceived the emotionally expressive animated agent as an important element of the computer mediated environment because it used emotion to provide subtle cues that highlighted important content in the lecture. The learners also perceived this agent as friendly and professional looking. However, there were some negative perceptions regarding the fidelity of the agent. The learners did not perceive this agent as human because it lacked proper synchronization between the topic of the dialogue and the expression of emotion (i.e., wrong emotion for the context). Another negative perception regarding the low emotion intensity animated agent was that learners felt it distracted them from focusing on the audio/voice component of the lecture.

Similar to the low emotion intensity agent, the learners that interacted with the high emotion intensity agent also perceived the agent as an important element to the instructional environment because the emotion contributed to their character as instructors of a lecture (similar to instructors in real life). On the other hand, there were positive and negative perceptions regarding the enjoyment of interaction. While some learners enjoyed the emotion and movement of the agent, others felt the agent was not visually pleasant. The perceptions of humanness of the high emotion intensity agent were mostly negative because of glitches in the synchronization of mouth movement and the audio. Finally, the learners in this treatment also perceived the

agent as distracting to the content of the lecture due to the facial emotion expression.

Even though, these findings mixed (i.e., positive and negative perceptions), they serve as a good source of design recommendations that could aid the design and development of agents in future projects. The first design recommendation is to utilize animated agents in circumstances in which humans would not be able to perform the same task or they could not perform the task with the same level of detail as an agent. For example, it is good to utilize animated agents in a computer mediated environment in which the agent has to perform a task that could be dangerous for humans (e.g., explaining how to deactivate an explosive). By doing so, the designer of the instruction highlights why it is important to utilize animated agents in that specific context. Another recommendation is to design the agent in a manner that it is visually pleasant to the learner. This can be done by utilizing an agent that the learner can relate to (i.e., relate in age or physical appearance) It might also be important to utilize an agent that the learners would find physically attractive and that seems relevant to the tasks they are performing in the computer-mediated environment. One more recommendation, specifically related to emotion representation, is to develop the agent in a manner that there it presents a good synchronization between the dialogue of the agent, the movement of the mouth, the motional inflection of the voice, and the emotional facial movement. This will prevent the learner from getting distracted and it will also further enhance the level of fidelity of the agent. One final design recommendation is to present facial expression of emotion in a manner that is as natural and subtle as the emotion express in humans, with emotional patterns that vary significantly throughout the interaction with the learner.

9.0 CONCLUSION

To summarize, this qualitative research found that participants in both treatments, low and high intensity emotion, perceived emotionally expressive animated pedagogical agents in a similar manner. Nonetheless, there are few limitations in this study. One of the major limitations of this study is the number of participants. In future research, a larger sample of participants should be recruited. In addition, the topic presented by the animated agent was one that focused mainly on negative aspects of African history. This content limited the facial and verbal expressions of the agent to mostly negative emotions (i.e., angry, sad, scared). It did not allow the implementation of a larger sample of emotional expression. Future researchers of animated agents should consider a topic that has a broader range of emotions (i.e., happy, surprise, disgusted, angry, sad, scared).

REFERENCES

- [1] Craig, S.D., Gholson, B. & Driscoll, D.M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2), 428-434.
- [2] Johnson, L., Shaw, E. & Ganeshan, R. (1998). Pedagogical agent on the web. Available: <http://www.isi.edu/isd/ADE/papers/its98/ITS98-WWW.htm>
- [3] Adcock, A.B., Duggan, M.H., Nelson, E.K., & Nickel, C. (2006). Teaching effective helping skills at a distance: The development of project CATHIE. *The Quarterly Review of Distance Education*, 7(4), 349-360.
- [4] Romero, E.J. & Watson, G.S. (2010). Emotion in animated pedagogical agents performing as virtual social actors. *Proceedings of the Fourth Modeling, Simulation & Visualization Student Capstone Conference 2010*. Suffolk, VA.
- [5] Atkinson, R.K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94(2), 416-427.
- [6] Baylor, A.L & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Journal of Education Computing Research*, 28(4), 373-394.
- [7] Bickmore, T.W. & Picard, R.W. (2005). Establishing and maintaining long-term human computer relationships. *ACM Transaction of Computer-Human Interaction*, 12(2), 293-327.
- [8] Dirkin, K.H., Mishra, P., & Altermatt, E. (2005). All or nothing: Levels of sociability of a pedagogical software agent and its impact on student perceptions and learning. *Journal of Educational Multimedia and Hypermedia*, 14(2), 113-127.
- [9] Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A. & Bhogal, R.S. (1997). The persona effect: Affective impact of animated pedagogical agents. *Proceedings of 1997 Conference on Human Factors in Computing Systems*.
- [10] Veletsianos, G. & Miller, C. (2008). Conversing with pedagogical agents: A phenomenological exploration of interacting with digital entities. *British Journal of Educational Technology*, 39 (6), 969 – 986.
- [11] Kim, Y., Baylor, A.L. & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23, 220-234.
- [12] Veletsianos, G. (2009). The impact and implications of virtual character expressiveness on learning and agent – learner interactions. *Journal of Computer Assisted Learning*, 25, 345 – 357.
- [13] Wertz, F.J. (2005). Phenomenology research methods for counseling psychology. *Journal of Counseling Psychology*, 52(2), 167 – 177.
- [14] Patton, M.Q. (2002). *Qualitative research and evaluation methods* (3rd ed.), Thousand Oaks, CA: Sage.

5.0 SERIOUS GAMES & VIRTUAL WORLDS TRACK

Track Co-Chairs – Talib Hussain and Mike Thome, BBN Technologies



5.1	Mission At Mubasi - A Simulation For Leadership Development	339
5.2	A Storm's Approach; Hurricane Shelter Training in a Digital Age	348
5.3	Sensing and Virtual Worlds – A Survey of Research Opportunities	357
5.4	Implementing Artificial Intelligence Behaviors in a Virtual World	367
5.5	The Impact of Incentivizing the Use of Feedback on Learning and Performance in Educational Videogames	374
5.6	Why Games Work and the Science of Learning.....	383
5.7	Virtual Worlds; Real Learning: Design Principles for Engaging Immersive Environments	393
5.8	Metaverse Utilization in the EFL Classroom - A Study. Second Life® as Part of an Institution's Curriculum.....	403
5.9	Crossing the Virtual World Barrier with OpenAvatar	410
5.10	Visual Tool for Authoring Learning in Games (VITAL)	417
5.11	Analog Design for Digital Deployment of a Serious Leadership Game	428
5.12	Zero-Fidelity Simulation: Engaging Team Coordination without Physical, Functional, or Psychological Re-Creation	435
5.13	Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions	444
5.14	Innovating Training through Immersive Environments: Generation Y, Exploratory Learning, and Serious Games	452
5.15	The Role of Game Based Learning in the Health Literacy of African American Adolescent Males.....	460
5.16	What Research Says About Competition and Learning	467

5.1 Mission At Mubasi - A Simulation For Leadership Development

Mission At Mubasi - A Simulation For Leadership Development

Paul Cummings, Steven Aude

Jon Fallesen

ICF International

Center for Army Leadership

pcummings@icfi.com, saude@icfi.com

jon.fallesen@us.army.mil

Abstract: The United States Army is investing in simulations as a way of providing practice for leader decision making. Such simulations, grounded in lessons learned from deployment experienced leaders, place less experienced and more junior leaders in challenging situations they might soon be confronted with. And given increased demands on the Army to become more efficient, while maintaining acceptable levels of mission readiness, simulations offer a cost effective complement to live field training. So too, the design parameters of such a simulation can be made to reinforce specific behavior responses which teach leaders known theory and application of effective (and ineffective) decision making. With this in mind, the Center for Army Leadership (CAL) determined that decision-making was of critical importance. Specifically, the following aspects of decision-making were viewed as particularly important for today's Army leaders: 1) Decision dilemmas, in the form of equally appealing or equally unappealing choices, such that there is no clear "right" or "wrong" choice 2) Making decisions with incomplete or ambiguous information, and 3) Predicting and experiencing second- and third-order consequences of decisions. It is decision making in such a setting or environment that Army leaders are increasingly confronted with given the full spectrum of military operations they must be prepared for. This paper details the approach and development of this decision making simulation.

1.0 INTRODUCTION

The ambiguous, less than perfect information and limitations of human perception, match well with the contemporary operating environment which US Army leaders routinely face. For this reason the Center for Army Leadership (CAL) determined decision making in such a situation was a high priority developmental need for its leaders. Consequently, resources were dedicated to a simulation that would provide leaders with practice in this type of decision making. A first step in the simulations development was to identify appropriate decision making theories, principles and models around which the simulation would be developed. Naturalistic decision making was identified early on as a way of thinking about decision making consistent with what Army leaders actually experience. A naturalistic framework emphasizes real world decision making in demanding situations that require cognitively complex thinking skills. From among the various models that take a naturalistic approach, two in particular were important to the development of the simulation.

2.0 THEORETICAL APPROACH

Snowden and Boone's (2007) Cynefin Framework (See Figure 1), the first approach for decision-making, provided a useful way to think about the context for the simulation. The framework identifies five contexts, defined by the predominant cause and effect relationship found in each, in which a leader may find him- or herself: simple, complicated, complex, chaotic, and in-disorder. In keeping with the important aspects of decision-making we identified, our focus was on the ordered complicated and unordered complex domains. Complex contexts are characterized by the lack of a clearly right answer, unpredictability, and flux. Snowden and Boone argue that in such a context, a leader should first probe or investigate, then assess the facts, and finally respond. Such contexts often require creative or innovative decisions, which may be accompanied by a risk of failure. Consequently, inexperienced leaders or leaders who are not comfortable with the uncertainty of the situation may fall back to a more controlling style of decision-making.

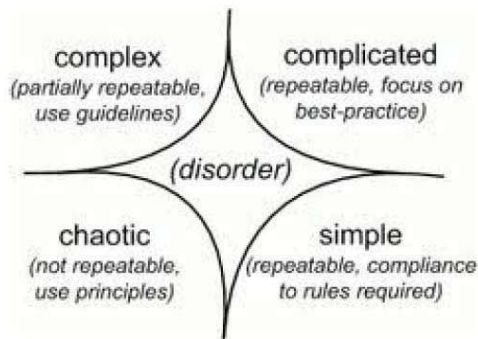


Figure 1: Cynefin Framework

The second theoretical model that provided structure to the decisions featured in the simulation was Quinn and Rohrbaugh's (1981) competing values framework. The primary tenet of the model is that organizational effectiveness is dependent on meeting numerous performance criteria that are organized by four value sets. At its core, this model is based on tensions – tensions between values and tensions between choices. In applying the competing values framework, we attempted to create tension between decision options. One source of tension was that for many of the scenarios, the decision choices were equally attractive or equally unattractive. There were numerous trade-offs among the choices, so that no single option was clearly advantageous. Another source of tension was between immediate outcomes and long-term outcomes of the decisions. For example, for a particular choice, the immediate outcomes may be very positive but the long-term outcomes are disastrous for one or more parties. Alternatively, another choice may lead to negative immediate outcomes, but the long-term outcomes are beneficial. An added advantage to using the competing values framework was that it contributed to the realism of the decisions. Often, real-life decisions are not clear-cut and trade-offs must be weighed.

Combining these two naturalistic approaches, the Mission at Mubasi simulation puts leaders in leadership dilemma situations which are initially represented as simple and ordered choices. As the simulation progresses choices become unordered and complex.

The learner must gather data from multiple sources and become aware of patterns in the environment in order to evaluate a best course of action. Thereby the simulation progresses towards more ambiguity and the learner must examine problems beyond single variable possibilities.

3.0 TECHNICAL APPROACH



Figure 2: Technical Framework

Our initial objective was to create a simulation that allowed the learner to evaluate competing information and determine a best course of action. It was decided that, in order to constrain complexity, the framework would be designed to present decision dilemmas as discrete points in the simulation rather than continuous actions taken to shape the simulated environment. These discrete decision points would be multiple choice dialogue interactions where virtual avatars (one or more) would provide context to the decision.

4.0 DEVELOPING MISSION AT MUBASI

An implementation of Mission at Mubasi was created for the US Center for Army Leadership's Multi-Source Assessment and Feedback (MSAF) program called Mission at Mubasi. The context of the simulation is to enable senior NCOs and junior officers to enhance their decision-making skills by immersing them in an engaging interactive game based training environment.

Mission at Mubasi is based on a fictitious North African country that is in need of governance and infrastructure development. Within the game the learner plays a leader new to the African country situation (i.e.

taking over from a previous leader) and the environment is negative for several reasons including lack of support for the Army governance, lack of ample food and water, and minimal security. There are also multiple parties with differing, sometimes conflicting, priorities and interests that the leader must address. In the course of the simulation, the leader goes through a series of scenarios in which he or she is provided with incomplete, ambiguous information from multiple sources, and then must select an option from various courses of action, none of which were designed to be ideal responses over the others.

Game Variables: Within the game the learner's objective is to maintain four important aspects of the governance process: relationship and positive affect of the villagers, well being of his soldiers, security in the environment, and transitioning governing capacity to the local tribes. Each of these values will be modified during the simulation by decisions and actions taken by the learner.

Conditions will arise either through proactive decisions by the user, or injects (*events*) that will dynamically modify the conditions within the simulation. The Primary Variable is composed of several Micro Variables (MVs). These MVs are assigned a weight of importance by the developer, and are modified directly by the simulation in order to determine the value of the Primary Variable.

Resources: Within the simulation the learner is responsible for optimally using and maintaining several resources towards the success of the mission. These resources include food, money, political capital, medical supplies.

Win Game Scenario: If the user is properly engaged in the simulation and is generating positive results, the Primary Variables (villager relationship, soldier well-being, etc.) should all be increasing towards a *success threshold*. This threshold value for each of these variables will be built into the simulation where once all the variables have surpassed this value, the simulation is complete. The communicated reward for the leader is seeing all of the positive results in the scenario. Conversely, if all four variables fall below the minimum threshold, and remain there for a designated period of time, the leader is advised that the situation has deteriorated to the point where the scenario is unwinnable. They may restart the game and try again.

5.0 DEVELOPING SCENARIOS

Our approach to developing the scenarios was iterative and systematic, while leaving room for creativity and emotional verve. In all, 11 scenarios were developed. Initially, we developed a single, prototypical scenario to practice applying the theoretical principles discussed previously and to refine our process. Our first step was to write brief descriptions, about a paragraph in length, for each scenario. This allowed us to create an overarching storyline across all of the scenarios, bringing cohesion to the scenarios. These descriptions also served as a map for developing each of the scenarios, such that when developing an earlier scenario we could create decision options that would set up later scenarios. Within each scenario there were several details including dialogues between participants, immediate, long-term effects, multi-order decision impacts, resource modifications, and decision points for the learner. Table 1 below contains a short summation of the scenarios and competing values taking place in the simulation.

Table 1: Example scenarios and competing values

Scenario	Summary	Competing Values
Meeting with Village Elders	The overall objective is to make decisions about the initial allocation of resources. The leader is new to this assignment and this is the leader's first meeting with all the elders together. This is a critical meeting for the leader because it will set the tone for his/her relationship with the local population. In addition, because the leader is new to this assignment, subordinate Soldiers are getting to know him/her as well. Consequences of decisions in this scenario will have long-term impacts because the relationships are just forming.	• Ensuring there is adequate security and food for both villagers and refugees • Desire to make villagers happy and build a positive relationship vs. limited resources to allocate toward village • Building relationships with the leader of the refugee camp and the village
First Attack	The leader must determine the threat level against the Forward Operating Base (FOB) and decide how best to ensure its security while continuing to perform governance tasks in the village. Ultimately the leader must determine how many Soldiers to assign to security tasks and what type of security tasks they should be performing.	• Assigning Soldiers to security tasks within the safety of the FOB; • Assign some Soldiers to ambush patrols outside the FOB versus allocate Soldiers to build infrastructure for the village.
Captured Son	Your Soldiers have just gotten in from an all-night patrol and are putting away their equipment. During the patrol, there was an ambush and subsequent firefight. One of your Soldiers was shot in the back. The medic does not know if he will walk again. Your Soldiers capture one of the teenage assailants who says he was coerced by the rebels into taking part in the ambush. However, one of your Soldiers is sure that the teenager had a weapon and was firing it. As your Soldiers are putting away equipment, one of the local leaders stops by to tell you that you have mistakenly taken his son prisoner and demands his immediate release. The local leader promises you that his son will never take part in a rebel army attack on US Forces again. It is the perfect opportunity to build goodwill with a local leader and have him in your debt, yet the very idea of letting go a potential rebel Army recruit go could send the wrong message. And your Soldiers would see you as too soft on security and too keen on "building goodwill" at their expense. Do you release the son or not?	• Soldier morale vs. relationship building with the local leader • Taking a firm stance on security vs. building goodwill
Security Dilemma	The village leader is walking purposely through the village on his way to a meeting with the local police chief. The local police force is at about 50% of its strength but they were stripped of their weapons and equipment when the rebels initially swept through the area. He has requested that you re-arm his police force with weapons so that they can once again fulfill their role in maintaining law and order in the village. Yet some of the killed rebels you came across were partially clad in police uniforms and bearing police side arms.	• Security of the village vs. rebuilding local policing capacity • Increasing security vs. building a relationship with the villagers
Hiring Dilemma	The village Water & Sanitation Director was killed when the rebels initially attacked the village and a new one must be hired. The council of elders would like to be in charge of finding the new Director, but you are not sure how quickly they can do this or whether expertise and experience will be the most important criterion used for selecting a new Director. Clean water and sanitation are key to the health of the village and preventing outbreaks of disease that could threaten everyone. The council assures you that they will find someone very good, but it doesn't seem like the council has established a good process for working out disagreements (at least without outside help), and your observation has been that key positions are given out based more on relationships and tribal status than actual expertise. Consequently you are very concerned with how they will decide on someone to fill a high-level post.	• Building a positive relationship with the villagers vs. building capacity of the village.

Farmer Loyalty	You are given a situation report at the FOB that there is trouble brewing between your Soldiers and the local farmers. You decided to drive out to one of the farms, where a group of farmers has gathered to meet with you. A group of your Soldiers has been working with local farmers to try to get the farms back up and running and enhance irrigation systems in the area. While they are working with the farmers, they start realizing that the farmers are selling grain and food to the rebels. When they confronted the farmers about it, the farmers finally admit that the rebels are paying much more for the grain than they are able to get by selling supplies to the refugee camp aid workers, who you had thought were the major buyers. The farmers are also fearful of retribution if they refuse.	• Building a relationship with the villagers vs. increasing security (by not allowing the farmers to supply the rebels)
Refugee Crisis	An NGO aid worker and their coalition security force say a new wave of refugees just showed up. She says apparently there is not enough shelter, medical supplies, food and water for all the refugees who have just arrived. Some riots and fights have broken out inside the fenced perimeter of the camp. Refugees are fighting over the limited shelter space and place in line for food and water. When you get there, you can see the aid workers trying to distribute the limited food supplies with panicked looks on their faces. They have now moved their truck outside the wire, tossing food bags over the fence to the fighting mob. Technically you do not have responsibility for the refugee camp, and the villagers are opposed to giving up resources for the refugees, but the refugees could present a threat to the NGO workers and Mubasi if they start foraging for food.	• Gaining political capital with the village vs. responding to the humanitarian crisis at the refugee camp. • Gaining political capital with the village vs. intelligence gained from helping the refugees.
Sick Cattle	Cattle are getting sick at one of the biggest farms. It turns out they have been poisoned by toxic weeds leaves that have been spread around the grazing area. A significant number in the herd will need to be put down. Many are convinced that it's the Lansi tribe members in the town (sympathetic to the rebels) who are at fault. The local police chief (who is Lansi) say he will lead the investigation. Local Mutsi elder objects – saying that the Lansi police chief is biased. The Mutsi want the U.S. to lead the investigation. However, if you don't let the chief do the investigation, you will clearly be undermining his authority, and the villagers may go back to depending on you to settle disputes. Who do you put in charge of this investigation?	• Building local capabilities vs. building a relationship with the villagers
Flash Flood	A flood did major damage to the water treatment plant. It also destroyed the bridge between the farmland to the east and Mubasi, virtually cutting the route by which farmers take their crops to market, and processed foodstuffs and supplies come to Mubasi. How will you allocate your resources to ensure safe drinking water, food, without compromising security?	• Relationship with the villagers vs. security • Relationship with the villagers vs. Soldier morale
Hiring Dilemma Part 2	Sanitation has been a disaster (evidenced by trash everywhere and a growing rat population). Your senior medical personnel tell you that the village is at high risk for an infectious disease outbreak if the trash is not picked up and the rat population continues to grow. Soldiers could contract various diseases, as well. Furthermore, you just found out that many of the staff at the Sanitation Department have not yet been paid, but the money seems to have all been spent. The council is very nervous about removing the sanitation director because his family is very powerful and he is very popular. He insists that trash is not picked up and people have not been paid because his department is woefully underfunded. The council will not recommend that he be replaced. What do you do?	• Building a relationship with the villagers vs. building infrastructure • Soldier well-being vs. relationship with the villagers



Figure 3: Mission at Mubasi Environment

6.0 VIRTUAL ENVIRONMENT AND GAMEPLAY

The Mission at Mubasi Decision-Making Simulation was developed using a 3D immersive web deployable gaming engine accredited for use on the Army Knowledge Online Portal. The virtual environment contained several indoor and outdoor areas (refugee camp, downtown, water treatment plant, FOB, farm, school house, and new construction areas) and over 400 virtual avatars that provided interaction and feedback to the learner while immersed in the gaming environment. As the learner progressed through the simulation, visual and audio cues were provided to help guide the progress on each of the primary variables. For instance if villager well-being was low, villagers would sit idle in the streets and less activity would be taking place in the market.

7.0 BETA TEST

A beta test of the Mission at Mubasi Decision-Making Simulation was held at an Army classroom with 16 participants in the range of senior non commissioned officers, and junior officers. Participants were given

introductory information on how to use the simulation, and were then asked to play through eleven scenarios and present objective and subjective feedback responses to the Mission at Mubasi simulation.

- *Will this lesson be useful to you on the job? If so, how? If not, why not?*
- *Were the decision dilemmas significantly challenging?*
- *Was it clear that decisions you made were tied to outcomes (1st and 2nd order effects) in the virtual world?*
- *What aspects of the simulation did you find engaging or stood out as well done?*
- *What, if anything, would you change about the simulation?*
- *Please provide any additional comments you have about the lesson.*

8.0 RESULTS

Beta test results were evaluated and categorized into two general categories, positive and negative impacts of simulation elements. Although there were many comments about the overall gameplay, Table 2 below presents the general responses we received from our two beta test groups.

Table 2: Responses from Beta-Test

Simulation Element	Positive Results	Negative Results
Understanding Competing Values (Quinn Model)	Learners felt that the dilemmas presented were generally representative of their deployment experiences.	Needed more information to determine best course of action (COA) Needed better understanding of story and mission to determine best COA.
Understanding Complexity and Decision Making (Cynefin Model)	Decisions were easily understood based on feedback within the simulation (visual cues, GUI elements) After Action Review helpful to understand impacts	At times difficult to find patterns of behavior within simulation
Engagement	Experiences engaging and sometimes fun	At times experiences were too complex to determine best approach
Developing Strategies	Interested in developing strategies	Discrete outcomes minimized strategies
Visual/Audio Cues	Fun and engaging, strong affect in certain scenes (i.e. <i>Attack at the Forward Operating Base</i>)	Issues with realism of soldier uniforms.
Understanding After Action Review (AAR)	Straight forward information provided with both graphical and text content describing the decision and impact	Sometimes hard to understand multi-order impacts over time (concept can be mentally challenging). AAR didn't go into enough detail about complicated topics.
Navigation	Simple game based interaction with engaging visuals and audio cues	For non-gamers, sometimes difficult to understand 3D navigation and game User Interfaces. No Undo feature

Competing Value Model: In terms of responding to the decision dilemmas, several participants felt that without enough information, especially in the early phases of the simulation, it was difficult to evaluate a best course of action. But the consensus was that it is unlikely a leader would always have all relevant information available to them in a real-life situation, and sometimes 'you had to do the best you could'.

Multi-Order Impacts: Our second objective was to make sure that outcomes in the environment (multi-order effects) were clearly tied to decisions or groups of decisions. This was a difficult challenge as information in the simulation was coming from multiple sources; it was therefore decided that directing the learner towards specific cues would be very important in making sure they understood relationship between cause and effect.

9.0 CONCLUSION

One of the most challenging aspects of developing a complex decision making simulation is to make a simulation real enough to simulate the variation in decision impacts, but simple enough where the learner is not lost in a myriad dynamics taking place in the game. The Mission at Mubasi simulation will continue to be evaluated over the course of the next several months for its learning efficacy. There is a great deal to understand about how decision-making can be trained including a) which are the most appropriate decision making models?, b) how important and to what level of fidelity should decisions be modeled within a simulation?, and c) what metrics are necessary to determine how well the learner was able to understand decision making impacts?

A great deal of work within the simulation was designed to challenge the user to recognize cues in the environment beyond traditional analysis of cause and effect. The emergent patterns (unordered domain) were designed to suggest to the learner that multiple decisions that were weighted towards any one of the variables could jeopardize the end result of the game. For example, if the learner were to make well-being or security, eventually this would cause events in the simulation that would decrease the overall success of the mission. Data collected from our users showed that emergent pattern recognition was most helpful when cues were ample, relevant, and consistent in the simulation. If cues were occurring in simulation that did not appear to coincide with decisions, it was more difficult for the learner to recognize the decision impact. The instructional designers used user interface cues to help clarify multi-order effects. For example, animating energy bars on the user interface helped to focus the attention of the learner on global changes in the environment, rather than direct impacts based on immediate decisions.

Future versions of the decision-making simulation will incorporate a schema format for extracting metrics for evaluation within the After Action Review (AAR) process. Additionally, it is our intention to continue to refine tools for experts to incorporate decision-making exemplars into the simulation framework.

10.0 ACKNOWLEDGMENTS

We would like to thank the US Center for Army leadership in support of our simulation development for the Multi-Source Assessment and Feedback Program. We would also like to thank Jorge Salazar, and August "Chip" Beyer for active participation in the development of Mission at Mubasi.

11.0 REFERENCES

- [1] Arkes, H. R. & Hammond, K. R. (Eds.), (1986). *Judgment and decision making: An interdisciplinary reader*, New York, NY, US: Cambridge University Press, xiv, 818 pp.
- [2] Brunswik, Egon (1956): Perception and the representative design of psychological experiments (2nd ed., rev. & enl.), Berkeley: University of California Press.
- [3] Fiske, S. T. and Taylor, S. E. (1984). *Social Cognition*. New York: Random House.
- [4] Flin, R., Strub, M., Salas, E. & Martin, L. (Eds.) (1997) *Decision-Making under Stress. Emerging Themes and Applications*. Aldershot: Avebury.
- [5] Klein, G. (2003). *Intuition at work: Why developing your gut instincts will make you better at what you do*. New York: Doubleday.
- [6] Orasanu, J., & Connelly, T. (1993). The reinvention of decision making. *Decision making in action: Models and methods*.
- [7] Orasanu, Judith; Connolly, Terry. In Klein, Gary A. (Ed); Orasanu, Judith

- (Ed); Calderwood, Roberta (Ed); Zsombok, Caroline E. (Ed), (1993). *Decision making in action: Models and methods*, (pp. 3-20). Westport, CT, US: Ablex Publishing, xi, 480 pp.
- [8] Quinn, R.E. & Rohrbaugh, J. (1981). A competing values approach to organizational effectiveness. *Public Productivity Review*, 5, 122-140.
- [9] Shaban, R. Z. (2005). Theories of judgment and decision-making: A review of the theoretical literature. *Journal of Emergency Primary Health Care*, Vol. 3, Issue 1-2.
- [10] Snowden, D.J. & Boone, M.E. (November, 2007). A leader's framework for decision making. *Harvard Business Review*, 59-69.
- [11] Ward, E. E. (1954). The theory of decision making. *Psychological Bulletin*. 51:380 – 417.
- [12] Ward, E.E, & Tversky, A. (eds) (1967). *Decision-Making: Selected Readings*. Penguin Books

5.2 A Storm's Approach; Hurricane Shelter Training in a Digital Age

A Storm's Approach: Hurricane Shelter Training in a Digital Age

Andrew Boyarsky
CUNY School of Professional Studies
aboyarsky@gc.cuny.edu

David Burden
Daden Limited
david.burden@daden.co.uk

Anders Gronstedt
Gronstedt Group, Inc.
anders@gronstedtgroup.com

Andrew Jinman
Daden Limited
andrew.jinman@daden.co.uk

Abstract: New York City's Office of Emergency Management (OEM) originally ran hundreds of classroom based courses, where they brought together civil servants to learn how to run a Hurricane Shelter (HS). This approach was found to be costly, time consuming and lacked any sense of an impending disaster and need for emergency response.

In partnership with the City of New York University School of Professional Studies, Gronstedt Group and Daden Limited, the OEM wanted to create a simulation that overcame these issues, providing users with a more immersive and realistic approach at a lower cost. The HS simulation was built in the virtual world Second Life (SL). Virtual worlds are a genre of online communities that often take the form of a computer-based simulated environments, through which users can interact with one another and use or create objects. Using this technology allowed managers to apply their knowledge in both classroom and remote learning environments. The shelter simulation is operational 24/7, guiding users through a 4 ½ hour narrative from start to finish.

This paper will describe the rationale for the project, the technical approach taken – particularly the use of a web based authoring tool to create and manage the immersive simulation, and the results from operational use.

1.0 INTRODUCTION.

The New York City Hurricane Shelter Simulation is a 3D immersive Hurricane Shelter training course developed in Second Life. The environment was modelled on a real School building in Brooklyn civil servants across the country have been using this simulation to up-skill on how to open, operate and close a hurricane center in an emergency situation. During this 4 ½ hour simulation users are guided through a structured narrative around the anticipated approach of a Hurricane Leo. Users work in groups through a total of 62 exercises created using PIVOTE (PIVOTE is an open-source authoring system for learning in virtual worlds, created by Daden Limited).

The PIVOTE system controls all the exercise logic and is and managed using several simple web interfaces.

The project earned a second place award in the AI (Artificial Intelligence) Assisted Training category of the Federal Virtual Worlds Challenge from a pool of about four dozen entries. The purpose of this paper is to identify the reasoning behind using innovative technologies such as Virtual Worlds; we intend to describe the following and achieve an in-depth analysis of our findings;

- Investigate the reasoning behind using a simulation technology rather than traditional education methods.
- What PIVOTE is and how it assisted in the development and scalability of the project.
- How the project has been utilized and tested to date by NYC civil servants.
- The future roadmap of the project, including duplication and roll out.

- And present our findings and observations of learners using this innovative technology.

Our approach to training presents further research opportunities that are worth exploring as the project develops in the future.

2.0 BACKGROUND CONTEXT.

In 2007 CUNY SPS started the coastal storm training project, with the intention to train 32,000 staff / city workers. The largest group, with the greatest training requirements was found to be Hurricane Shelter Managers.

New York City hurricane shelter managers, numbering over 2,000, were required to take a full day of traditional instructor-led class-room training. While feedback from the class room sessions was positive, many participants stated that they wanted more hands on work in the facilities with the equipment and the situations that they would likely encounter in a large scale sheltering response. In response to this feedback, New York City Office of Emergency Management has successfully run live hurricane exercises at facilities designated as emergency shelters both in 2006 and 2008. However, the logistical and cost requirements of training 2,000 managers this way does not make it a scalable option. Instead, they opted to develop an immersive 3D program.

The four-hour-long learning simulation takes place in a virtual hurricane shelter developed in Second Life that is modeled on a real school building in Brooklyn. The learner is playing the role of a Hurricane Shelter Manager - opening, running, and closing the shelter. This 3D virtual rehearsal and training simulation helps City employees practice their shelter management skills and apply problem-solving skills in a realistic environment. They can take the self-paced simulation on their own or in a small team of three to five players. "Our feedback suggests that the cooperative experience of playing and being scored as a small group is more effective than a single-player experience," says Andrew Boyarsky, Project Director at the School of Professional Studies, CUNY

Graduate Center, who spearheaded the program.

3.0 OVERVIEW OF THE SIMULATION.

The simulation starts with a newscast announcing that "Hurricane Leo" is forecasted to hit the New York City area in a few days and Hurricane Shelters are being activated. The learner has just arrived at the shelter from the Evacuation Center. They practice setting up the facility for shelter clients when they arrive, designate key shelter areas, assign staff, and ensure that areas are set up and supplied properly so the shelter is ready to open to the public. When the shelter clients arrive, learners are challenged to solve a number of thorny problems. How do they resolve conflicts? How do they communicate with guests who don't speak English? The AI recognizes key words and determines the next response by the non-player character. The complete simulation includes 40-50 such non-player characters driven by the AI system. Individual scores are captured in the PIVOTE web-based authoring system.

As the clock ticks down to zero hours, learners prepare for the arrival of the storm. After the storm, they conduct a walk-through to check that the building is safe, and address post storm concerns of the shelter clients and the needs of the facility. For some issues they will find solutions in a Field Guide while others will require some thought using managerial skills and experience. Learners need to think on their feet, just like in the real world. People learn in context, while earning points, solving problems and receiving feedback.

3.1 What is PIVOTE and why use it?

Before PIVOTE the standard way of creating learning exercises in the Second Life virtual world was to embed scripts written in Linden Scripting Language in the in-world objects. Text was either embedded in the scripts, or placed in "notecards" which were also embedded in the objects. This approach required a relatively high technical skill level, made maintenance a nightmare since to change a piece of text or behaviour you had to locate the relevant

object in-world and make the change, and meant that the whole exercise was bespoke to Second Life.

In developing a paramedic training system for St George's Hospital [1], London Daden were introduced to Medbiquitous Virtual Patient (MVP) [2], an XML (Extensible Markup Language) standard created to enable medical schools across the globe to share "virtual patients" - simple training exercises for medical professionals based around a medical case and usually following a "branching" or "string-of-pearls" structure [3]. It was realised that the MVP standard could also be used to define a virtual world training exercise, allowing the exercise to be defined and maintained in MVP XML through a readily accessible and manageable web interface, but with the user experiencing the exercise within the virtual world by interacting with in-world objects and people. With this approach the scripts in world could be kept far more generic – and effectively just passed an object ID and a user action (e.g. chest – touched) back to the MVP "player" to work out the relevant response and pass that information back in-world. In this way all structure, logic and non-3D content of the exercise was kept in a manageable form on the web as MVP XML, and the bespoke in-world content was kept to a minimum. Added advantages of this approach was that every student action was automatically detected and could be easily time stamped logged, and the exercise could potentially be played from an environment other than the virtual world – such as the web.

The new MVP based system was called PIVOTE [4]. The MVP player, which manages the student interaction during an exercise, is written as a web-service. This allows any client environment to be potentially used as a PIVOTE front-end, from HTML (Hyper Text Markup Language) and Flash to mobiles.

There are several key reasons why a PIVOTE-based approach makes sense for virtual world exercise development:

- **Maintenance and scalability:** The exercises are created online rather

than in-world, making it easy to maintain and alter content without any SL or scripting expertise needed. The system is extremely robust and can be used by non technical members to create new exercises relatively quickly and with minimal technical input. The system is also multi-user allowing many people to play through the same exercise at any one time (of course the maximum number is dependent on land capacity of the virtual world technology).

- **Platform neutrality:** Since all the structure and non-3D content of the exercise is stored on the web the solution is independent of SL; meaning; the exercise will not be tied to this particular platform if they wished to port their exercises to another technology for any reason. These may include mobile technologies, flash, or any other virtual world.
- **Accessibility:** PIVOTE provides a platform where if learners do not have access to a Second Life capable PC they can still participate in the scenarios via a simple web interface.
- **Reporting:** PIVOTE records all interactions that take place in world between objects and learner for analysis online at a later date, including time-stamping of every participant interaction with the system, and allows export of this data for later analysis.

By implementing a PIVOTE based solution for this project:

- We can use the code already in PIVOTE to track participant activity without having to create a bespoke solution for this project at additional cost.
- We can use the object and animation control software already in PIVOTE to implement and manage the majority of the exercise – again without having to write new code and again saving cost.
- We can use the HUD (Heads up display) management used on past PIVOTE exercises – again without

having to write new code and again saving cost.

- The structure of the exercise will be more clearly visible and changeable through the web – rather than spread through multiple scripts in multiple objects, reducing on-going support costs.
- All of the non-3D content can be managed and edited through the web – allowing rapid implementation of simple text changes.
- Participants can be given access to a text-based version of the exercises.

4.0 SIMULATION COMPONENTS & GAMIFICATION.

There are flaws in the way we use traditional education techniques, users lack engagement and are demotivated by having to take part in assessments. When a learner for example is being formally assessed through a standardized test or similar device they instantly become anxious to do well and alter their behavior; this can sometimes have an effect on the outcome. This reaction is referred to in human behavioral science as the Hawthorne effect, when those who are being observed react to the knowledge that they are being observed [5].

With the HS Simulation project we wanted to employ these new techniques and avoid some of the traditional pit-falls. The project was designed to be fun, engaging, realistic, collaborative and most importantly educational. By using a technology like Second Life we opted to use a number of Gamification techniques throughout the simulation trainer. "Gamification can be explained as the process of game thinking and use of game mechanics to engage audiences and solve problems." [6] For this project the important thing we aimed to achieve was "Gamifying" the practical hands on experience learners would have in a real emergency situation, making the training more exciting and re-engaging than with traditional methods.

Using and applying a number of gamification techniques we aimed to encourage the learners through the methodology - "progress encourages

progress". To do this the user is presented with clear, tangible objectives in the form of exercises and modules which give a sense of flow. We believe the human desire for efficiency and to do well acts as a powerful motivator here as the users gain instant feedback on their decisions. This is supported by the sense that the user is embodied in the Role of the Hurricane Shelter Manager and is not simply being tested for it.

These were achieved by using a number of key components inspired by commercial game based techniques. In the following section we will describe these components and define why they strengthen the learning experience throughout the training simulation.

4.1 School and Environment.



Image 1 The school building

The school and Environment were modeled on a real school building and surrounding area in Brooklyn, as shown in image 1. Providing a backdrop and content for the learner to relate to. Many schools and public buildings across America are deemed Emergency Centers in case of an emergency and by providing such an environment for this simulation it provided a sense of realism and grounding.

4.2 Narrative and Video.

Throughout the simulation the story of Hurricane Leo is told; the storm is reported from a variety of characters; including an instructor, shelter clients and most importantly a fictional news report. This is given at the beginning of each module and in the form of reward for completing the previous module.

The news reports provide a sense of participation in the event, conveying both

what is happening and providing a sense of context and urgency.

The news reporter uses a number of graphical aids (weather charts / simulations of Hurricane trajectory) and fabricated interviews with specialists, with the solo purpose of immersing the learner in the story and current situation.

4.3 Characters.

There are four types of characters (primatars) found throughout the simulation, characters in the exercises, which are interactive and talk to the user during the course. HS staff and police are found throughout the simulation, these represent the staff and security that would actually be present during the use of the Hurricane Shelter. And finally shelter clients, these add a sense of realism and busyness around the shelter. Shelter clients also make statements in the chat channel as you walk past them. Example statements include -

[05:21] Shelter client: I hope my snake is OK...

[05:22] Shelter client: My cot is surprisingly comfortable...

Again by employing this simple technique it adds to the learners "sense of realism" and adds to the overall narrative.

4.4 Animations.



Image 2 An example animation

Animations are found throughout the simulation – these are specific animations triggered when the user accomplishes an exercise. They see the result of their action played out in the 3D world. These animations are developed in-world and triggered via PIVOTE, an example of an animation can be seen in image 2. After the learner has answered a question regarding

unpacking supplies; staff form a daisy chain to bring the supplies into the building.

4.5 Exercises.

An available exercise is identified by the red question mark above the point of interaction. To continue and begin the exercise a user needs to click the question mark. This system is traditionally seen in more commercial games, for instance in World of Warcraft, available quests are indicated by different symbols.

In the simulation a question mark has three states, red for available exercise; amber for exercise in progress and green for exercise complete. This helped indicate to the user at which stage they are in an exercise and their progress. We found this gave users a sense of satisfaction and achievement by providing visual feedback and encouragement as they progressed. It also helped to reassure them they were using the system correctly.

For ease of use the exercises throughout the simulation follow a particular rule; PIVOTE was programmed to display the following sequences of text, when a user enters and exits an exercise, we found this format was preferred by the test groups; The only disadvantage to this system was the dependency on local and server side lag, which may cause the text to display out of sequence. Image 3 shows an example of the conversation sequencing.

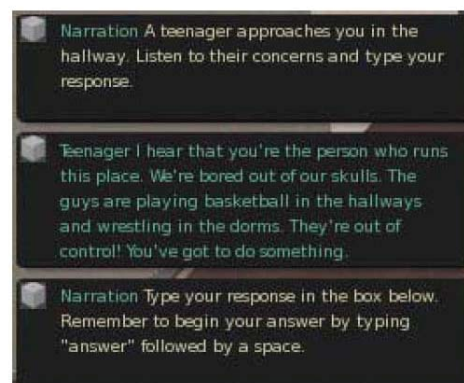


Image 3 Example conversation

4.5.1 Starting an exercise

1. **Description HUD text:** i.e. "A teenager approaches you in the..." (Yellow text).

2. **Primatar dialogue:** i.e. "Teenager I hear that you're the person who runs this place.." (green text).
3. **Primatar audio transcript** (associated with primatar)
4. **Narration HUD instructions:** i.e. "Type your response in the box below..." (Yellow text).

4.5.2 Ending an exercise

1. **Primatar dialogue:** i.e. "Teenager That is sooooo cool! I'll..." (Green text).
2. **Final audio transcript** (associated with primatar)
3. **Narration HUD feedback** i.e. "You should have found some places..." (Yellow text).
4. **Narration HUD instructions** i.e. "You've completed this activity. You have <number> more to go to complete this assignment. Follow the arrows to the next activity and then click on the red question mark." (Yellow text).

4.6 The Narration HUD.

The HUD is the learner interface for the simulation, it lists their score and allows them to access help resources. Two important aspects the score and the storm count down, both of which provide the learner with instant feedback on their progress through the simulation.

4.7 Navigation.

Arrows are located throughout the HS simulation to show the user where they need to go next. The arrows are only displayed once an activity is completed, connecting and displaying a path between exercises. This addition was added after the initial user tests; the school building was found to be too large and learners found it hard to navigate around the course without getting lost or taking much longer to find the exercises. This visual assistance has to be extremely helpful allowing learners to move seamlessly between exercises without delay.

Again the arrow system also reinforces the learner's sense of achievement, unlocking the next path as they progress, from exercise to exercise.

4.8 Storm system.

The storm system is made up of a number of components. The Hurricane, The Rain system, Sensor system, Hurricane Audio, and random rubbish blowing in the wind.

Together these elements provide a sense that a Hurricane is approaching. These components are controlled using PIVOTE and based on the user's behavior in world. The storm begins as the learner is instructed to head to the command center; when entering the building a sensor is triggered and storm begins.

5.0 USER EXPERIENCE.

The course was rolled out at the start of 2011 with an early Beta test of 7 sessions. A total of 14 participants took part in the course with an average of 2.14 participants per session. With over 93% responding to a course evaluation, the following quantitative results were tabulated on a five point (Liker) scale 5-strongly agree, 4-agree, 3-neutral, 2-Disagree, and 1 Strongly Disagree. The results were as follows:

Rating Area	% Indicating they Strongly Agree or Agree with the Statement
Did the program make you more confident in participating in the coastal storm plan sheltering system	86%
Did the program include training activities that were effective	86%
Did the program provide training in Hurricane Shelter activities that were closer to real world activities than classroom based courses.	100%
Did the program engage you as a learner	100%

Table 1. Evaluation results.

The first two ratings above are comparable to other types of online courses in the

curricula. However, the second two ratings are significantly higher than those for other online course formats. Comparing the last rating for engagement, at 100% vs 87% for a 1.5 hour flash based course with similar content, this is a difference of 13 percentage points. Indeed, when informally polled at the end of the course whether they would take the course again if given the chance, an overwhelming majority agreed they would.

These evaluations also capture open ended comments and the feedback is largely positive. Some of the most frequent comments from participants:

- The course was interactive.
- Provided more hands on work where and with what you would be dealing.
- More real life, allowing visualization

Experienced shelter managers and emergency responders have given the simulation high acclaim, expressing how great it is to have activities in a realistic environment, providing an experience as close to the high fidelity of a real life exercise.

In fact, the course administrators have observed that with groups of 3-5 course participants they quickly form a team with one person assuming the role, without prompting or direction, of shelter manager with the others taking direction. This team formation facilitates social learning between peers as one participant will refer to directions from the embedded field guide that should be followed. Further to this point, we have observed a greater sense of team cohesion. One participant was observed telling a caller on his cell phone that he needed to get back a team exercise and would call later on a break, the first time something like this was witnessed in overall shelter training program. These aspects are not as present in both classroom or in individual 2D (flash based courses).

Often at the start of sessions, it was observed that participants will remark that they do not engage with these types of technology. Typical participant comments are along the lines of “my kids play these types of games”. Or “That is not for me”. When they start off they are often

psychologically “on the fence”, and within an hour they appear to have psychologically “stepped off the fence” and are fully engaged. By the end of the simulation when the project director has entered the room to tell them the course is close to over participants commonly do not notice he is in the room they are so engrossed and hardly realize that four hours have passed!

6.0 DISCUSSION.

CUNY SPS in partnership with the NYC Office of Emergency Management is looking to expand the use of the simulation in a couple of ways. The course will be replicated on three additional islands (server spaces that can host an independent instance) in Second Life. While this allows for a greater number to partake in the simulation (4 islands x 5= 20 per session), it will also allow the simulation to be used in a larger exercise with multiple facilities; a cluster of hurricane shelters will be managed by an evacuation centre, allowing NYC OEM to test command and control within the hierarchy.

Within NYC, many City agencies use City schools (Dept. of Education) for emergency response plans, e.g. Points of Dispensing (PODs) for mass vaccinations, commodity distribution post emergency (C-PODs). NYC OEM is considering adding minor enhancements and to coordinate the use of the simulation for these multiple plans as it reduces the expense of holding exercises and allows for the immersive training that has been previously described.

NYC OEM is also part of the Regional Catastrophic Planning Team that coordinates emergency preparedness planning and response for catastrophic disasters. This includes four states and 30 counties in the New York/New Jersey metro area, 22 million people or 1/14 US Citizens. NYC OEM is considering expanding the use of this platform for emergency shelter training and exercises for multiple levels for its regional partners. This simulation could possibly be targeted to a potential regional audience of 150,000 shelter staff.

CUNY SPS would like to evaluate the effectiveness of this medium for both training and exercises in contrast to those in both classroom and 2D (flash based) courses by assessing different cohorts with similar characteristics in actual emergency shelter deployments.

6.1 PIVOTE.

In looking at how to move PIVOTE forward in the future we have focused on two areas – the ease of use of the exercise authoring tool, and extending PIVOTE to control the layout of the 3D space as well as the logic and content of the exercise. In both of these we have been informed by some of the prior work within the field of eDrama by an organisation called Hi8us.

The design of the authoring screens in PIVOTE reflect the data hierarchy in MVP - Activity Nodes , Availability Nodes, Asset Nodes. Whilst this architecture gives significant potential flexibility it is not something we have ever needed in practice and currently serves only to confuse the author. MVP was also a very “link” based approach – nodes automatically lead to other nodes, whereas a virtual world environment is more “real” - student choices are what determines what node follows each node. What we have found is that clients and authors are far more comfortable with the eDrama terms of “prop” and “actor” (an actor being a non-player-character in the scene). The new version of PIVOTE will use a props and actor metaphors, with each prop 9 and actor) having a set of associated behaviours based on the possible student actions. We think that this will make the system not only more usable, but also significantly enhance re-use of objects between exercises.

Whilst PIVOTE abstracts the logic and non-3D content of an exercise into XML on the web, it still needs a 3D specialist to create the 3D environment. Having moved to an eDrama approach it makes sense to bring the positioning of props and actors into PIVOTE. The intention is that the author is presented with a set of libraries for different scene types (e.g. city emergency, rural

emergency, hospital, industrial site etc) and uses a 2D web-based drag and drop environment to create the scene (including elements like floors/roads, building walls and even terrain backdrops). Then either through the layout interface or a simple prop list the allowable actions and behaviours of each prop and actor can be defined – even specifying routes for actors and mobile props. Once the scene is complete – and tested on the web – the user can then save it, and issue the command to have it instantiated (called “rezzing”) in the virtual world.

The next generation of the PIVOTE system is currently in development – called OPAL – in conjunction with a UK Academy School – with the aim of making it usable by teachers and students as well as by training professionals.

7.0 CONCLUSION.

The driving force for using a virtual world platform such as Second Life was to provide a hands on experience for assigned emergency shelter managers in the working environment (a City school as an emergency shelter) performing the activities using the equipment and other materials that would be available during an emergency sheltering response. This need emanated from previous training season feedback by assigned shelter managers. The emergency shelter managers assigned are mostly City school facility managers, and, while they are familiar with the building they work in and their normal operations, they are not familiar with the operation of an emergency shelter. Until the release of the Hurricane Shelter Simulation the training available for shelter managers has largely been classroom based and comprised of table top exercises that incorporate various multimedia to help simulate the environment. Unfortunately, as well designed as the classroom exercises are, they fall short of providing the level of fidelity of the simulation.

The simulation provides a sense of self-relevance and context for the learner through the use of the fictional news reports, game aspects, and audio and chat input. Prompting the learner to move about the facility, to undertake the various tasks

and problem solving, gives the individual and team a sense of responsibility and ownership that is not available in both a classroom and 2D (e.g. flash based) courses.

This approach to online, simulation based training suggests that the gap between training / role-playing in the classroom can be bridged with the use of technology. As the project develops we aim to continue a comparative study of the costs involved for development and our original findings of the live hurricane exercises run during 2006 and 2008. Indication already suggests that scalability makes a simulated approach more cost effective compared to traditional classroom based training.

Using a technology of this type isn't without its challenges and limitations; One consistent observation is the mind-set of the learner and their apprehension to use such an innovative platform. Even though during our trials users quickly overcome these apprehensions in groups. It is imperative that we consistently consider the typically "less technical" user demographic during our development, and ensure we make every concession needed to make the HS Simulation easy to use.

8.0 REFERENCES.

[1] Coradi E and Burden D et al. (2009). Virtual patients in a Virtual World: Training Paramedic students for Practice. *Medical Teacher*, Volume 31, Issue 8, pages 713 – 720

[2] Ellaway, R., Candler, C., Greene, P. and Smothers, V.. (2006). *An Architectural Model for MedBiquitous Virtual Patients*. Baltimore, MD: MedBiquitous

[3] Poulton T. et al. (2009). The Replacement of 'Paper' Cases by Interactive Online Virtual Patients in Problem-Based Learning.

[4] Burden D et al. (2008). Creating and Assessing a Virtual Patient Player in Second Life. *Researching Learning in Virtual Environments Conference*. Open University.

[5] McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P (2007). "The Hawthorne Effect: a randomised, controlled trial"

[6] Gabe Zichermann, Christopher Cunningham, 2011; *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*.

5.3 Sensing and Virtual Worlds – A Survey of Research Opportunities

Sensing and Virtual Worlds – A Survey of Research Opportunities

Dana Moore
SAIC
dana.moore@saic.com

Abstract: Virtual Worlds (VWs) have been used effectively in live and constructive military training. An area that remains fertile ground for exploration and a new vision involves integrating various traditional and now non-traditional sensors into virtual worlds. In this paper, we will assert that the benefits of this integration are several. First, we maintain that virtual worlds offer improved sensor deployment planning through improved visualization and stimulation of the model, using geo-specific terrain and structure. Secondly, we assert that VWs enhance the mission rehearsal process, and that using a mix of live avatars, non-player characters, and live sensor feeds (e.g. real time meteorology) can help visualization of the area of operations. Finally, tactical operations are improved via better collaboration and integration of real world sensing capabilities, and in most situations, 3D VWs improve the state of the art over current “dots on a map” 2D geospatial visualization. However, several capability gaps preclude a fuller realization of this vision. In this paper, we identify many of these gaps and suggest research directions

1.0 INTRODUCTION

Virtual worlds can add value to many domains. It is possible to draw many Venn diagrams that feature virtual worlds serving as the presentation or interaction layer for a wide variety of use cases. In this paper, we suggest that by combining live sensing, three-dimensional (3D) virtual worlds and social networking capabilities, we can derive a command and control capability that significantly improves communication, situation awareness, and situation response. Figure 1 depicts this conjunction graphically (C4ISR = command, control, communications, computers, intelligence, surveillance, and reconnaissance).

2.0 AN ILLUSTRATIVE USE CASE

Live tactical sensing field trials and exercises have become extremely expensive to stage and conduct, and even when the logistics and system integration go without a hitch, the after-action “hot washes” are often unrevealing—results are often hard to characterize, replay and data mine sufficiently well to arrive at meaningful conclusions. Often, the quality of logistics preparation is blamed for many of the problems in exercise setup, and the lack of customer focus on strong system integration. Both are commonly cited as



Figure 1: Intersection of ISR and virtual worlds

the cause of often indifferent and hard-to-interpret results.

We contend that while these both may be true, they are aspects of execution of a specific strategy that relies exclusively on live execution as its mechanism.

A different approach, one we will propose here, should be obvious to the modeling and simulation community: Use the real devices to stimulate the model and provide a high-fidelity simulation, high enough in many cases to render the actual exercise optional or even unnecessary. In this way, travel and logistics become moot, and many more exercises can be run within the same time and funding constraints. In fact,

it may be possible to use simulation for command and control of the actual in situ as a deployed system. Nowhere would the benefits be more keenly felt than in large, logistically complex exercises. We take as an illustrative use case the National Level Exercise, yearly mandated by the White House. While this exercise is pure simulation, not involving real, deployed sensors, it does serve to illustrate the aspects of exercise complexity and the utility of virtual environments, and lay the foundation for later assertions we will make.

3.0 NATIONAL LEVEL EXERCISE

Each year, the U.S. executive branch mandates an exercise called the National Level Exercise (NLE)[1], which is done only as a virtual exercise. One reason given for this “virtual-only” approach is that the dynamics of the exercise, which models the kinetic event, is of sufficient magnitude that putting the exercise together live would be prohibitively expensive and time-consuming. According to the U.S. government’s own literature, the NLE is a part of the “*National Exercise Program (NEP), which serves as the nation’s overarching exercise program for planning, organizing, conducting and evaluating national level exercises. The NEP was established to provide the U.S. government, at all levels, exercise opportunities to prepare for catastrophic crises ranging from terrorism to natural disasters.*”

The 2011 NLE concluded in May simulated a 7.7 magnitude earthquake on the New Madrid fault near Memphis, TN. The geo-specific terrain in the simulated environment encompassed 10 square kilometers with over 2,100 structures at sub-meter photo-real high resolution.



Figure 2: NLE 11. High overhead view of downtown Memphis, TN

Figure 1 shows a screen capture of the 3D virtual environment in which NLE was implemented.

Pre-exercise, the commercial On-line Interactive Virtual Environment (OLIVE) was used for over four months as a collaboration, exercise development and systems integration environment. During the exercise, the 3D environment was used as the command and control nexus for 14 command groups whose participation included several participants, including U.S. state, U.S. federal, U.S. Army Northern Command (NORTHCOM), and U.S. Coast Guard participants. A control room with participants in remote, real-world locations, but co-located in virtual space, and several simulated system feeds, again from disparate sources in multiple data clouds, can be seen in Figure 3.

By any standards of assessment, NLEs have been well received, produced significant data, and by design, exercised all aspects of emergency response from low-level tactical to the highest echelons of state and federal government that we assert might not be possible to otherwise conduct.



Figure 3: NLE 11. Command and control center with multiple simulated data feeds

For purposes of this discussion, we can consider NLE as an example of large-scale integration, mission preparation and command and control, all staged and performed in a virtual arena. If actual sensors had been used to stimulate the virtual arena rather than synthetic data, the architecture could support a live exercise.

With NLE as a reference implementation, we now have a frame of reference in which we can examine the question of how one might move from a fully virtual arena to a real-world, operational domain. This examination will help characterize technology gaps that have, in the past, prevented technologists from achieving integration of 3D simulations of areas of interest with operational sensing capabilities. These gaps, in turn, represent opportunities for research. It is important to note that significant gaps exist both in the state of the art in sensing and in 3D virtual worlds that attempt to mirror real-world attributes (which often are called “*mirror worlds*”).

To attempt to bound what is meant by “sensors,” we suggest restricting our

purview to those that have proven relevant in exercises such as NLE and would be commonly used in a tactical setting.

Amongst these would be

1. Environmental data, such as temperature, precipitation, wind speed and direction, clouds and visibility, day/night (based on actual ephemeris models), wave heights and riverine flows
2. Unattended ground sensor/unmanned aircraft system (UGS/UAS) data from traditional sources such as infrared, seismic and acoustic sources; magnetics and thermal sensors
3. Video from electro-optical sources, such as unmanned aerial vehicle (UAV) feeds, webcams, closed-circuit television (CCTV), broadcast television
4. News (and other) data from streaming sources, such as Really Simple Syndication (RSS) feeds, broadcast monitoring systems, language capture and translation

To these, we might add a new and interesting, but non-traditional, abstract sensor that might provide insights into the human terrain of an area of interest (AOI) and might, for example, drive models of

non-player characters. These might include:

5. Social/cultural indicators, providing real-time, geo-specific indicators of social *climate* or *temperature* based on real-time news feeds
6. Social networking, useful for estimation of the memetic ecology of an AOI. These might be instant messaging; Twitter[®] (Twitter, Inc.) tweets; email

We begin the discussion by looking at sensing and then move to mirror worlds.

4.0 RESEARCH PURSUITS FOR MODERN SENSOR SYSTEMS

A readily apparent characteristic of sensor systems is that while they are capable of producing copious data, they often yield little actionable data absent much massaging, either by a human in the loop or by automated systems. It is therefore insufficient, we contend, to simply faithfully model a sensor and show some manifestation of its data outputs in a mirror world. Thus, although one approach might be to faithfully model a weathervane, dropping it into a virtual mirror world as shown in Figure 4, the tactical value is very small. There is added value in seeing the sensor *in situ*, with the surrounding environment fully navigable in 3D.



Figure 4: Wind direction and velocity sensor model fed by real-world source (Harvard School of Public Health CitySense Project)

However, while having multiple perspectives might be useful and can add situation understanding, the data from the sensor itself has a limited utility, constrained by the timeliness of the reported data. Were the sensor less “dumb” it might be able to yield some notion of its operating data (e.g., its remaining duty cycle, or its error envelope), or trend data, long term or short term, or the inter-relationship with other environmental physical sensors.

One can argue that long-term persistence of such data and meta-data, or converting it into useful intelligence, is the responsibility of systems seeking to integrate sensor data, but therein lies the root of a perennial problem. Traditional sensor systems are at once too “dumb”—often producing too much data, and yet too little actionable intelligence. A sensor field might incessantly report the footsteps of a few goats on a hillside, but be unaware of its mission context, which might be to report not goats, but men traveling in formation.

This aspect of the sensor field’s mission might be well understood by humans receiving the data, but because of the incessant reporting, might come to be ignored as noise source, especially problematic when the real formation of men comes across the hillside.

In part, this situation exists because traditional sensor systems, especially physics-based sensors, have been designed, implemented, and fielded as a part of a bespoke system. Such systems are heavily “*silo-ed*,” useful in very narrowly defined contexts, with interfaces dictated by the size constraints, communication and power budgets of traditional micro-electronics components. The value of modeling such a sensor field in a 3D environment would be essentially nil, but would entail considerable software development cost.

However, owing to a virtuous trend of micro-componentizing compute and communications capability, functionality narrowly constrained by package size or power limitations need no longer be the case. Piece parts such as the *Gumstix*[®] [2] (Gumstix, Inc.) embeddable computer, running accustom Linux[®] (Linus Torvalds) stack are readily available. 3GS and WiFi piece parts are likewise easily available. Were these simple piece parts a commonplace in physical sensors, a paradigm shift would shortly occur.

Thus, we can identify the first of many research pursuits: **Add standardized local computation, and communications capability to sensors.** Doing this by creating a common sensor reasoning and communications platform and creating standard interfaces would benefit an entire creative community, but especially the 3D command and control virtual tactical operations center.

One may argue that abstract, non-physics-based sensors (e.g., RSS feeds) do not suffer the same lack of local computation capability. Although this is essentially correct, the current state of the art is that even for networked abstract sensors, while there may be compute resources attached, these are seldom used to add value, to change data into intelligence.

Hence, our second suggested research pursuit: **Add local intelligence, reasoning and context understanding to sensors.** Reasoners (e.g., C Language Integrated Production System, CLIPS [3]) can be packaged and deployed in quite small packages and placed adjacent to individual sensors or as a “super node” in sensor fields, but the challenge is to develop better reasoning capabilities beyond venerable reasoning suites founded on Bayesian-style logic.

The need for better sensor reasoning is predicated on the belief that a common sensor vernacular or ontological

representation can exist. Currently none does, although one may argue that SensorML [4] could be the foundation of such a data description language or, at the very least, a useful conceptual antecedent.

In any case, it does point out the potential for a research pursuit: **Develop sensor agnostic description languages that might become the basis for query and description.** Innovators in the 3D virtual environment would use description and query languages to place sensors automatically in the 3D scene graph and query them for timely update; show information flows among sensors and command and control paths; and automatically add and operate non-player characters or other automata into the scene graph.

Another capability that would dovetail nicely is the addition of marketplace mechanics to sensors or sensor fields. This was the intent of the short-lived and fondly remembered JXTA[®] (Oracle America, Inc.) [5] protocol and framework, wherein small form-factor data producers (e.g., sensors) offered their content into a marketplace and market mechanics played a strong part in steering the mission. The JXTA architecture was at once elegant, comprehensive and compact. A research pursuit derived from this observation: **Develop a marketplace mechanic for sensors and webs of sensors.** A marketplace mechanic would enhance the capability to create a virtual operations center (as well as enabling the capability to more easily control sensor fields from mobile devices).

Interwoven with representation are issues of access. As mentioned previously, accessing and extracting data from sensors has been traditionally difficult. Currently, most virtual worlds can interface with almost anything that responds to an Extensible Markup Language Hypertext Transfer Protocol (XMLHTTP) Request [6] or XML remote procedure call (RCP)

protocol [7] requests and responses. Real-world sensors are not amongst the data sources that respond to these protocols.

This is unfortunate, since there is general agreement that Web and rich Internet application (“cloud”) frameworks have redefined popular thinking about how distributed applications are imagined, architected and implemented. This suggests yet another research pursuit: **Create a capability to add small footprint, agile, language-based web frameworks** (e.g., Pylons [8], Ruby on Rails[®] (David Heinemeier Hansson) [9], TurboGears [10]) **to sensors and sensor fields to enable them expose their data both as web servers and web clients.** While there are open source sensor web architectures for composition of large-scale sensor networks such as 52North [15], many consider this an extremely heavyweight approach.

The modeling and simulation community would especially benefit if the goal of incorporation of real-world inputs is ever to be realized. Not having to write custom interfaces for each sensor type would represent major forward progress. Yet, although these tools have been available for almost a decade, little thought has been applied to using web technologies in sensing.

Agile languages can play another part in putting real-world sensors in play in virtual worlds. Most onboard sensor algorithms are still coded in low-level languages such as Verilog[®] (Gateway Design Automation Corporation) [11], or even cross-compiled C. In contrast, most virtual worlds use agile (sometimes referred to as “scripting”) languages to provide the interaction points in their environment. Bigworld[®] (Bigworld Pty Limited), a popular game engine uses Python [12]; Unreal[®] (Epic Games, Inc.), another popular engine, uses UnrealScript, a Java[®] (Oracle America, Inc.) variant; Unity 3D[®] (Unity Technologies APS), a hugely popular game and virtual world

engine uses JavaScript[®] (Oracle America, Inc.) and C#; and finally Second Life[®] (Linden Research, Inc.), a well-regarded virtual world engine uses its own C-family-derived compiled scripting Linden Scripting Language (LSL).

It is clear that there is a gap here—there is no useful mutual language that facilitates a common expression of algorithms and business logic in a lightweight, highly expressive language (e.g., Python[®] [Python Software Foundation], PHP, Lua[™] [Lua Technologies, LLC.] or Ruby[®] [Faculdades Catolicas]). One may argue that as long as the ability to support web paradigms such as Representational State Transfer (RESTful) [13] interactions amongst participants in a sensors and virtual worlds mashup, there is no compelling need for common languages.

Yet, the evolution of Java, and hence of object-oriented programming, was predicated on the belief that a single compute language could facilitate common expression across many platforms, scaling from supercomputers down to the very small. A mantra of the 1990s became “Write once—run anywhere,” and while this proved to be a goal, there were seams and gaps in implementation detail, it still serves as a brilliant reminder of the art of the possible.

Thus, a next suggested research pursuit: **Create a common high-level language that spans the requirements of sensors, sensor fields, and virtual worlds.** Since both environments are inherently event-oriented, a lightweight language supporting concurrent programming (co-routines, or lightweight threads), callback registration, and a high level of abstraction would be ideal. Loads of Python and other lightweight languages exist for embedded Linux platforms, such as the aforementioned Gumstix single-board computer, so this suggested pursuit may be regarded as within the realm of the

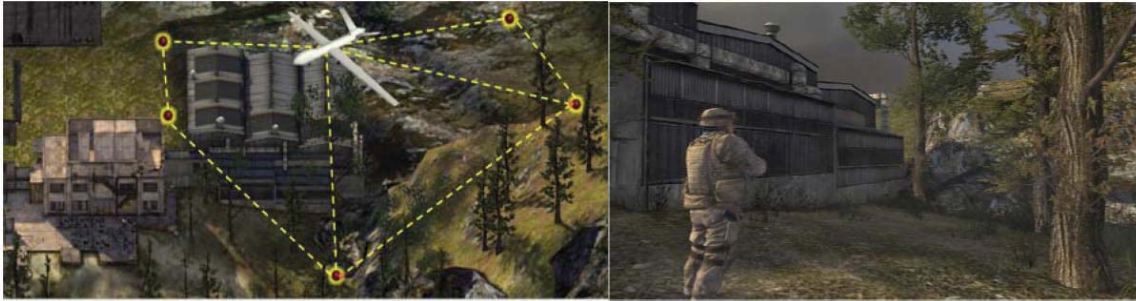


Figure 5: (A) Virtual world lookdown onto resources and AOI(B) A zoomed-in view over the soldier's shoulder

possible and not require a completely fresh conceptual start.

One may argue that had there had been a fully evolved Internet prior to the 1990s, the need for a common language would not have been as compelling or would be even a moot point, given a number of game engines and virtual world platforms that support some form of RESTful interaction. We would counter, however, that having a common object description approach can potentially lead to more efficient interaction and less impedance mismatch.

5.0 RESEARCH PURSUITS IN VIRTUAL WORLDS AND GAME PLATFORMS

There are many gaps in current sensing design and implementation that make a useful integration with virtual worlds more difficult and prevent engineers from achieving the kind of useful integration that creates the tactical operations center of the future. Figure 5 suggests what a tactical officer's view of the AOI might look like in our vision of the future. Notice that resources are visible in their relationship with the AOI in both the high aerial view and the dismounted soldier's perspective, and further, zooming in to perceive greater detail is seamless. This is not dissimilar to what one might experience in a "Google® Earth" (Google, Inc.) view of an AOI, in which one can zoom in from a very high

altitude down to a greater level of detail; hence, one might ask, Why is it better and what does the virtual world contribute to the tactical perspective? There are several perspectives on this question, but let us address a few of them in this paper.

First, let us consider tactical displays based on 2D geospatial platforms (e.g., ArcGIS® [Environmental Systems Research Institute, Inc.]) as shown in Figure 6, or 2.5D visualizations (Google Earth, e.g.) as shown in Figure 7.

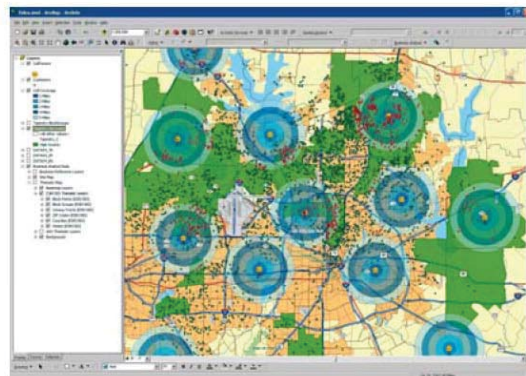


Figure 6: Typical Geographic Information System (GIS) display



Figure 7: Google Earth display augmented with "grey box" structure.

photographs. Like other 2D and 2.5D views, utility is significantly lower in the urban environment, and occlusions can deceive the user.

Figure 8: Video drone output

What research pursuits are important to bring the benefits of 3D navigation and immersive experience to the tactical user? Let us consider a few of these. First, the need for object interoperability has become extremely important. A number of model representations, including COLLABorative Design Activity (COLLADA) [16], 3DS Max[®] (Autodesk, Inc.) [17], FilmBox file format (FBX) [18], and control environments of long-standing, high-level architecture (HLA) [19], distributed interactive simulation (DIS) [20], virtual reality VR federates, currently exist. However, these are not easily composited into a game or virtual world environment. Thus, a first suggested research pursuit: **Create an object interoperability architecture that can enable models from different producers in a game- or virtual world-agnostic way.** Along with this, there would be an ability to distribute an arbitrarily large-scale AOI across multiple game platforms or virtual worlds concurrently.

The need for geo-specific terrain and structure or geo-typical terrain, decorated with specific structure is a critical success factor for convincing tactical users to move to 3D tactical viewers. Thus, a second



Figure 9: LIDAR point cloud example

suggested research pursuit for virtual worlds: **An ability for (semi) automatic generation of terrain and static structure.** Currently, 360° laser detection and ranging (LIDAR) can perform metrology down to the centimeter level, creating point clouds along with the texture for the point as shown in Figure 9; however, the ability of converting point clouds to meshes is still missing.

In addition to creating structure, there is a need for higher-fidelity graphics in virtual world platforms. The most often voiced complaint from military users when shown a prototype in a commercial virtual world such as Second Life is that it does not compare favorably with console or PC games. It does little good to point out that there is a difference between the open and unlimited vistas that a virtual world must support and the constrained scene graphs that a game engine can support. Thus, a third research pursuit: **Methods and algorithms to deliver higher graphic quality in virtual world engines.** Although Linden Labs are slowly introducing polygonal meshes into their renderers, they drag along a great deal of legacy code with them. Therefore, fresh approaches are needed to deliver higher quality.

Ironically, and at the same time, users clamor for lightweight clients, capable of running in web browsers or on mobile platforms. Hence, a final suggested research pursuit that will yield solid results: **Create virtual worlds platforms capable of being run in the browser, without need for plugins.** Currently, there is much

hope for HyperText Markup Language (HTML) 5 [21] as a standard that can host a virtual world within the browser, but much foundational work must be accomplished before this can become a reality.

6.0 CONCLUSIONS

As this paper has asserted, representing and using sensors in massively multiplayer online games (MMOGs) and virtual worlds can produce better tactical command and control, and improved situation understanding. Virtual worlds are an inherently better vehicle for full 3D navigation and, when driven by sensors in conjunction with competent artificial intelligence (AI) engines, for representing forces and counterforces in an AOI.

However, a long legacy of “doing things as they have always been done,” especially in the sensors world, limited their use in 3D contexts or in other advanced interfaces that incorporate augmented reality or full sensory immersion.

On the virtual worlds’ side of the equation, the trajectory of commercial virtual worlds such as Second Life or Teleplace® (Teleplace, Inc.) is such that they have only moderate interest in advancing the state of the art to address the gaps identified. In general, where topics are well aligned with their goals of attracting recreational users, one may expect to see progress benefitting sensor modeling and C2 coming from the virtual worlds community *as a side effect* of general advances in the state of the art. The general case is such that advances in visual representation come at a high investment cost on the part of the platform creator, and are of arguable value. Thus, the introduction of COLLADA meshes (to create more attractive avatars, clothing, and venues) is on the roadmap for Linden Labs and Second Life, but object interoperability is not.

Therefore, in both sensing and modeling disciplines there are a number of excellent opportunities that are “green field” for research exploitation.

7.0 REFERENCES

- [1] NLE Fact Sheet http://www.fema.gov/media/fact_sheets/nle2011_fs.shtm
- [2] gumstix: <http://gumstix.com/>
- [3] CLIPS Expert System Description: <http://en.wikipedia.org/wiki/CLIPS>
- [4] Sensor Modeling Language: <http://www.opengeospatial.org/standards/sensorml>
- [5] JXTA: <http://jxta.kenai.com/>
- [6] XMLHTTP History and specification: <http://en.wikipedia.org/wiki/XMLHttpRequest>
- [7] XMLRPC Home: <http://www.xmlrpc.com/>
- [8] Pylons Project Home: <https://www.pylonsproject.org/>
- [9] Ruby On Rails Project Home: <http://rubyonrails.org/>
- [10] TurboGears Web Framework: turbogears.org
- [11] Verilog for Electronic Design: <http://en.wikipedia.org/wiki/Verilog>
- [12] Python Language Home: python.org
- [13] Representational State Transfer: http://en.wikipedia.org/wiki/Representational_State_Transfer
- [14] HTML 5 High Level Description: <http://en.wikipedia.org/wiki/HTML5>
- [15] 52North: <http://52north.org/communities/sensorweb/>
- [16] https://collada.org/mediawiki/index.php/COLLADA_Digital_Asset_and_FX_Exchange_Schema and <http://en.wikipedia.org/wiki/COLLADA>
- [17] 3D Studio Max 3ds Format: http://en.wikipedia.org/wiki/Autodesk_3ds_Max
- [18] FilmBox File Format description: <http://en.wikipedia.org/wiki/FBX>
- [19] High Level (simulation) Architecture: [http://en.wikipedia.org/wiki/High_level_architecture_\(simulation\)](http://en.wikipedia.org/wiki/High_level_architecture_(simulation))
- [20] Distributed Interactive Simulation (DIS): http://en.wikipedia.org/wiki/Distributed_Interactive_Simulation
- [21] HTML Overview: <http://dev.w3.org/html5/spec/Overview.html>

5.4 Implementing Artificial Intelligence Behaviors in a Virtual World

Implementing Artificial Intelligence Behaviors in a Virtual World

Brian Krisler
Raytheon BBN Technologies
bkrisler@bbn.com

Michael Thome
Raytheon BBN Technologies
mthome@bbn.com

Abstract. In this paper, we will present a look at the current state of the art in human-computer interface technologies, including intelligent interactive agents, natural speech interaction and gestural based interfaces. We describe our use of these technologies to implement a cost effective, immersive experience on a public region in Second Life. We provision our Artificial Agents as a German Shepherd Dog avatar with an external rules engine controlling the behavior and movement. To interact with the avatar, we implemented a natural language and gesture system allowing the human avatars to use speech and physical gestures rather than interacting via a keyboard and mouse. The result is a system that allows multiple humans to interact naturally with AI avatars by playing games such as fetch with a flying disk and even practicing obedience exercises using voice and gesture, a natural seeming day in the park.

1.0 INTRODUCTION

When artificial intelligence (AI) entities exist in a virtual world, their existence is intended to fill out the scene and help to make the interaction more believable. When you walk down a street, you are typically not alone. There are other people around, in the shops and passing by in their vehicles. These individuals are all going about their personal activities, independent to you, yet they also have a role in your actions and goals. They can cut you off, slow you down or distract you from your intended goal. Should the need arise; you can also interact with these physical world agents using typical forms of communication such as speech and gesture. However in a virtual world, the extent of interaction allowed between humans and AI is often limited to a small vocabulary with an even smaller set of AI driven characters. These limitations produce an interaction that often feels shallow or fake - it is missing the level of realness provided by the *extras* encountered in the physical world.

Improving the realness of AI within a virtual world requires that the entities have the ability to act as entities within the physical world, where they seek and perform goals. To provide goal-seeking behavior to an AI



Figure 1: To demonstrate our architecture, we implemented a virtual dog in Second Life.

and have them exhibit the expected behaviors associated with the actions performed towards the attainment of those goals, we used an off-the-shelf rules engine to generate the behaviors of AI within a virtual environment. The rules engine provides the ability to assign actions and goals to an entity, where each action is based upon a set of facts. These facts are derived from the world and include the current state of the AI and any ongoing communications with other avatars. The occurrence of these internal and external events helps to drive the creation and attainment of goals, providing a level of

realism not typically seen in virtual worlds such as Second Life.

As a demonstration of our technology, we implemented a rules-driven dog avatar (**Error! Reference source not found.**) within Second Life. When in the virtual world, a user avatar can freely interact with our dog. The AI dog can perform the typical actions exhibited by dogs at any local park such as sit and play dead. To extend the level of realism, our dog also performs various idle behaviors. When left alone, the dog will wander about the scene, lie down or bark at random objects. When summoned, he will stop his idle behavior and come to the person calling him.

2.0 CURRENT STATE OF AI IN VIRTUAL WORLDS

A cursory review of AI within the history of gaming reveals two common threads:

- Tightly constrained interactions with the AI, as seen in NetHack [1], ELIZA [2], PARRY [3, 4] and Jabberwacky [5].
- Or rigid operational situations with no moral ambiguity and expectations that constrain the AI behaviors.

Each of these threads can produce a convincing AI within their well-defined context, however in a situation such as a human/AI conversation, their limitations become quickly apparent.

Within a virtual world, the AI (bots), while based upon the common threads, present relatively simple behaviors that are programmed with basic scripted responses or rely upon external engines to provide the intelligence. They are usually embodied as fancy virtual objects that are easily distinguished from human-operated avatars.

In Second Life, for instance, there are a vast number of objects, scripted with simple rule engines to provide primitive behaviors. These range from shop keepers, greeters, and bartenders, to autonomous animals and pets, to art installations that react to surrounding activity, but they all share the

feature in that their appearance diverges significantly from human avatars – they don't move, talk, or animate like people.

Another form of AI within a virtual world (VW) is an automated system that uses real avatars, essentially acting as expert systems controlling the avatar as a human operating a VW viewer would via a modified but otherwise standard viewer program or through a library that emulates such a program. However, these are most often used not to simulate intelligent beings in the world, but instead are used as surrogates for humans, allowing them to perform actions not allowed for scripted objects. For example, one of the first uses of these systems in Second Life was to automatically search the virtual world for cheap land for sale and snap it up. Another form of bot within a VW is a "model bot." Model bots have no AI per se, but rather use their embodiment to model clothing, and other looks realistically.

Finally, there have been a few, mostly dead-end, experiments in using embodied AI systems to interact with people on a peer basis. Generally, these have not been convincing and, as a result, have been relegated to research niches, and have never gained any widespread adoption.

3.0 OUR APPROACH TO AI IN A VIRTUAL WORLD

To represent AI within a virtual world, we modeled our approach on the way a human interacts with and communicates within the world around them. When a human confronts a world, either real or virtual, they are driven by a series of goals and actions. As they progress about the world and encounter others, the immediate goals can change, with new goals and actions constantly getting created, executed, and completed. For example, the simple act of walking down a street to meet a group of friends (Figure 3) could incur multiple sub-goals along the way. Each of these sub-goals consists of its own actions that are independent of the primary goal. The

insertion and execution of multiple sub-goals is an important part of modeling how a human operates with the world. Each new sub-goal operates without impacting the main goal. However in existing virtual environments, AI interact with the world independent of the actions around them. Because they contain a limited set of behaviors, they are restricted in the number of actions they can perform, preventing them from displaying new and unexpected

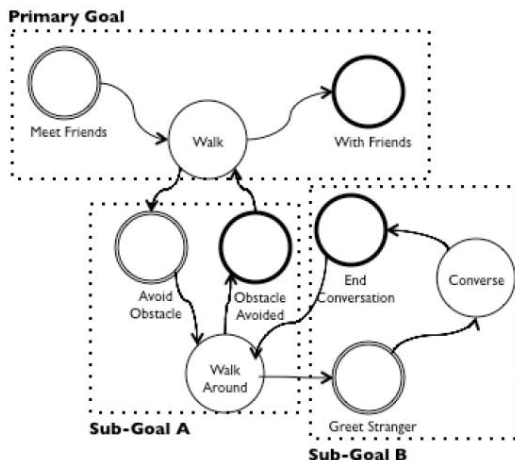


Figure 3. Goals consist of actions and sub-goals, where the insertion and execution of sub-goals occurs independent of the primary goal.

behaviors in unique situations, making their actions look scripted and artificial.

To improve the look and feel of AI in a virtual world, we approached the problem through improved input modalities and improved behavior modeling.

3.1 Improved Input Modalities

As part of our design, we wanted to create a more seamless interaction between the physical and virtual world. A critical component of this interaction is the ability to communicate with an artificial agent as seamlessly as communicating in the physical world. When a user interacts with a virtual world, they rely upon the keyboard and mouse to act as an interface between themselves and the world. However this interface provides an unnatural interaction, where all thoughts and actions must get conveyed using written text and graphical

interface widgets. This form of interaction does not easily lend to the interface disappearing [6], and misses the mark on user experience enrichment.

To attain a more seamless interaction, we extended the standard modalities used to interact with a virtual world. In our system, we added two new input methods: natural speech and physical gestures.

Through the incorporation of speech via a wireless headset and physical gestures using a time-of-flight camera [7], users are provided the means to interact with entities in the virtual world as they would interact in the physical world. Now instead of having to type *come here* into a console, a user can speak the phrase using natural language or wave their arm in a summoning motion to request the attention of a virtual avatar. If a user wants to play catch with their virtual dog, simply executing a throwing motion could result in an object getting thrown in the virtual world.

By providing the ability to speak real commands and interact using physical actions the *gulf of execution* [8] between the user and the virtual world is reduced, leading to a more enriching user experience.

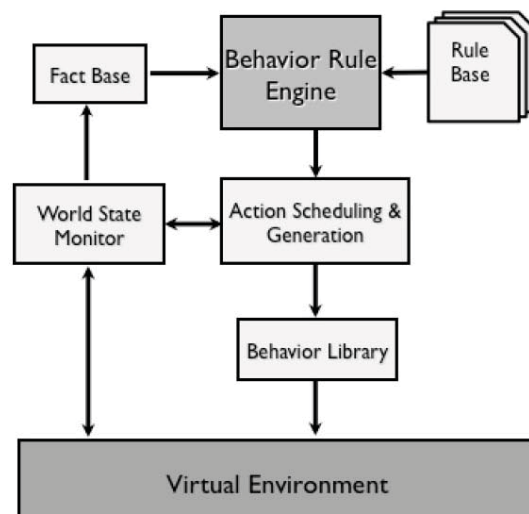


Figure 2. To create believable AI that exhibit goals and actions, we developed a behavior rules engine based upon facts and rules.

3.2 Improved Behavior Models

To produce a more realistic AI model where an agent can react to unexpected events, we based our design upon a behavior rules engine. A rules engine generates actions and behaviors in response to a combination of facts and rules (Figure 2) where facts characterize the current state of the agent and environment and the rules define the desired actions and behaviors that should occur for an observed factual state.

Each agent starts with an initial set of facts defining their basic characteristics such as name, age and gender. These initial facts set the stage for future actions and behaviors. When the agent enters a recognized world state, the facts of the world are used in conjunction with the initial facts about the agent to determine the appropriate action for the current situation.

Each of these fact-behavior definition pairs are defined as a rule within a rules base. Each rule consists of two parts: the conditional and the action. When the conditionals are met, the action is triggered. For example, an agent could have a rule for responding to a question, such as "What is your name?" where the conditional part of the rule consists of: *is the speaker looking at me and is the speaker within conversation distance*.

To determine the appropriate action, the rule engine utilizes the *Rete* algorithm [9]. The Rete algorithm is a pattern-matching algorithm commonly used in production systems to determine the best rule to fire for a given factual state. For example, an AI agent could approach the speaker's avatar in the virtual world in response to a request to *come here*.

4.0 IMPLEMENTATION

To implement our design, we combined a suite of open source tools with proprietary code and custom behavioral rules (Figure 4). These components were integrated the public, and unmodified Second Life virtual world in the middle.

4.1 The Artificial Intelligence

Interaction with the virtual world required a method for controlling avatars within the simulator. Typically, avatars within Second Life are controlled via a human operator. This operator communicates with the world through a client interface using inputs, such as a keyboard and mouse to manipulate the avatars behavior. However for our AI backed avatar, we had to replace the human-centric input mechanisms with generated behaviors.

To simulate a goals and actions based behavior, based upon the current state of the virtual world, we leveraged the open source rules engine *Drools* [10].

Drools is a Rete algorithm backed rules engine developed by the JBoss Community. Using *Drools*, we were able to create a set of rules to attain the behaviors we desired from our avatar.

In our implementation of a virtual dog, we required two sets of behaviors, *Active*

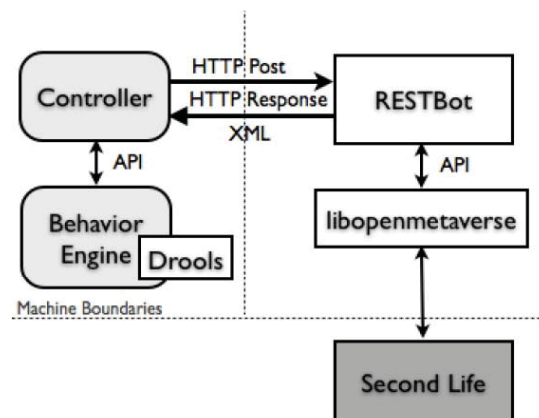


Figure 4. To implement the architecture, we relied on a suite of open source tools. All of the white boxes represent the open sources tools, and the light gray boxes are custom software. While the protocols support a distributed architecture, for our setup we ran all processes on a single computer.

Behaviors and Idle Behaviors. An Active Behavior is a behavior that occurs when the avatar is reacting to an input from the world, such as a request from another avatar. An Idle Behavior on the other hand is a

behavior that occurs during idle time, when there is no direct interaction with any other avatars. In our implementation, the idle behaviors include actions such as bark, lie down, and take a drink of water.

When the AI avatar is present in the virtual world (i.e. alive), the rules engine executes an idle behavior loop. This loop will generate random behaviors from a list of known idle behaviors, producing a dog that will wander about the scene, performing various actions. If another avatar interacts with the dog, the idle loop stack is cleared, allowing the active behavior to take precedence. Employing this approach causes the dog to stop whatever idle action it was performing and react to the other avatar. By defining a library of 10-15 actions, we were able to create an AI dog in Second Life that exhibited behaviors typical of a dog in the physical world.

To apply our actions and behaviors to an avatar in Second Life, we used the open source toolkit RESTBot [11, 12]. RESTBot is a REST based framework built on top of the open source toolkit, libopenmetaverse [13]. Running as a lightweight HTTP server, it listens for POST commands containing the desired interactions with the virtual world. When a command is received, it translates the HTTP message into virtual world actions. Upon processing of the command, any expected results are passed back to RESTBot, which in turn repackages the results in XML and passes the XML content back to the controller, all within a single HTTP transaction.

The remaining piece of the architecture was linking the Drools engine to RESTBot. To make this link, we implemented a simple controller based upon the command pattern that maps a generated behavior with a known RESTBot command.

4.2 The Human Interface

To reduce the interface between the user and the virtual world, we wanted to make the human operators' experience as natural as possible.

To attain a natural interface, we used a 3D TOF camera (a technology similar to Microsoft Kinect), coupled with software for tracking a users body position, resulting in the user's body becoming the controller [15].

The operational environment contains a defined physical area for the user to interact with the virtual world. When the system recognizes that a person has stepped into the scene, the user is instructed to stand still briefly while the user's body position and posture are recorded. The system then uses this information as a baseline to recognize future movements.

One of the control mechanisms implemented is the usage of moving in various directions to control avatar movement. When the user steps away from the neutral center position, it is recognized as a move in that direction, causing the avatar to move in same direction. Moving back to the original center position halts the avatar's movement. The avatar's directional movements mirror the user's movements. Step left to move the avatar left, back to move the avatar back, diagonally to move diagonally, etc.

Gestures, operator body postures and positions were combined with inputs from a Nintendo WiiMote to add some additional controls, all funneled into a modified full Second Life viewer application. The gesture system interacted with the viewer through another REST interface [14]

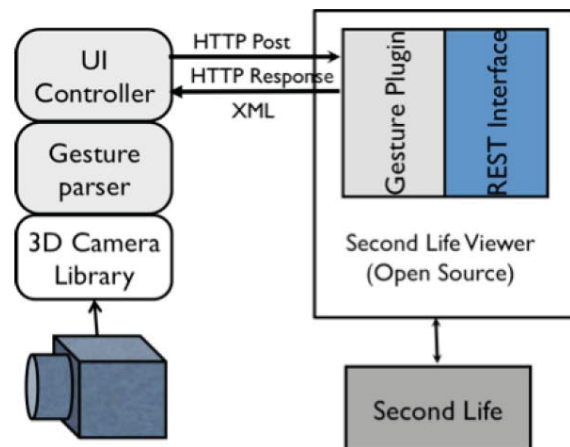


Figure 5: Architecture of User Interface

augmented with additional movement controls required to interact with our UI, controlling the avatar's actions much more directly and therefore less prone to error than having the controls merely sending keystrokes.

The viewer then displayed on a life size screen, putting the operator "into" the virtual environment and breaking another barrier to immersive experience.

4.3 Second Life Extensions

In order to implement a reasonably complex and immersive experience, we sourced most of the scene components from publically available vendors advertising in the Second Life Marketplace. The purchased products were assembled into a park space in one corner of a plot of virtual land we already owned.

We implemented a custom Frisbee-like game of catch for the purposes of this demonstration. While there are similar games available for purchase in Second Life already, none of the commercially available ones work well with non-human avatars, and we wanted our dog to be able to catch and carry the disc in its mouth rather than a paw.

Finally, there were a few additional minor pieces of supporting programming and building required, mainly to provide for a more seamless demonstration environment, improving overall performance, and adding a variety of sensors to improve gathering of metrics on the scenario.

To exhibit the behaviors expected from a dog, we created a few new commands for RESTBot. For example, we wanted the ability to have the dog *follow* a human controlled avatar. By adding in a new command to the RESTBot plugin library, we were able to quickly extend the tool to meet our requirements.

Once we had the ability to pass our generated behaviors and actions into the world, we hooked up our new forms of input to the behavior generator. Using BBN developed Speech recognition software we

were able to translate natural language into text that was then used by the rules engine to drive the behavior. For example, we created a rule that matched the work "speak" with an action that results in the dog barking.

We also developed a gesture library for recognizing a core set of gestures, such as an arm moving in a Frisbee throwing motion. Using this physical action as input to the rules engine, we were able to generate and *throw* a Frisbee in the virtual world that would then get fetched and returned by the dog.

5.0 FUTURE WORK

We would like to extend the temporal reach of the AI metaphor. In particular, the strength of non-player characters is that they are always "on" and always available to interact with. We intend to approach this goal in the following significant: use Second Life's Voice-over-IP system to communicate with the bots, and enhance the AI rule engine to have an attentional model to allow it to focus on a single human avatar at a time.

First, as a matter of available time and complexity, we implemented the voice recognition system to expect a direct input from a high-quality headset. Second Life does have a functional voice communication system, but achieving high quality voice recognition of low-quality audio is a notoriously difficult problem. Furthermore, the interface to the SL voice system itself is non-trivial.

Also, we plan to enhance the AI engine and rule set to be able to focus on a single human at a time when performing. A variety of distraction behaviors as well as complex modes of interaction with multiple humans and even other AIs becomes possible.

Together, these improvements would allow us to field highly interactive AI agents around the clock to interact purely through the commodity virtual world of Second Life.

6.0 CONCLUSION

By decreasing the interface between a user and a virtual world and improving the behaviors of AI within the world, the overall user experience can become more immersive, allowing the user to forget about the boundary between the physical and virtual world.

In our research, we have investigated an approach to decrease the interface through the removal of the keyboard and mouse barrier. We replaced these input modalities with natural speech and gesture providing a more natural interface to the world.

We also used off-the-shelf technologies to apply more realistic behaviors to an AI avatar within a virtual world. When used in combination with the more natural interface, a user can interact with the virtual agent as if it was another entity in the physical world.

7.0 ACKNOWLEDGEMENTS

The authors would like to thank David Diller, Rich Shapiro and Kerry Moffitt for all of their work on the design and development of the system.

8.0 BIBLIOGRAPHY

1. *NetHack* 3.4.3. [cited 2011 July]; Available from: <http://www.nethack.org>.
2. Weizenbaum, J., *Computer power and human reason : from judgement to calculation* 1977, San Francisco: W.H. Freeman.
3. *Robots (computer game)* - *Wikipedia, the free encyclopedia*. [cited 2011 July]; Available from: [http://en.wikipedia.org/wiki/Robots_\(computer_game\)](http://en.wikipedia.org/wiki/Robots_(computer_game)).
4. Güzeldere, G.a.F., S. *Dialogues with colorful Personalities of early AI*. 1995; Available from: <http://www.stanford.edu/group/SHR/4-2/text/dialogues.html>.
5. Carpenter, R. and J. Freeman, *Computing machinery and the individual: the personal Turing test*, 2009, Citeseer.
6. Abowd, G.D., E.D. Mynatt, and T. Rodden, *The Human Experience*. Pervasive Computing, 2002: p. 48-57.
7. *Time of Flight Camera*. August, 6, 2011 [cited 2011 July]; Available from: http://en.wikipedia.org/wiki/Time-of-flight_camera.
8. Norman, D., *The psychology of everyday things* 1988, New York: Basic Books.
9. Forgy, C.L., *Rete: A fast algorithm for the many pattern/many object pattern match problem* 1*. Artificial intelligence, 1982. 19(1): p. 17,Äi37.
10. *Drools*. [cited 2011 July]; Available from: <http://www.iboss.org/drools>.
11. *restbot - RESTbot allows bots in Second Life to be commanded through an HTTP interface - Google Project Hosting*. [cited 2011 July]; Available from: <http://code.google.com/p/restbot>.
12. Pleiades. *RESTbot Unveiled*. Available from: <http://pleiades.ca/2007/09/28/restbot-unveiled>.
13. *libopenmetaverse developer wiki*. May 5, 2010 [cited 2011 July]; Available from: http://lib.openmetaverse.org/wiki/Main_Page.
14. Sol, D. *User:Dzonatas Sol/SNOW-375 - Second Life Wiki*. [cited 2011 July]; Available from: http://wiki.secondlife.com/wiki/User:Dzonatas_Sol/SNOW-375.
15. Diller, D., et al. *Interacting Naturally in Virtual Environments*. I/ITSEC, 2010.

5.5 The Impact of Incentivizing the Use of Feedback on Learning and Performance in Educational Videogames

The Impact of Incentivizing the Use of Feedback on Learning and Performance in Educational Videogames

Girlye C. Delacruz

University of California, Los Angeles / National Center for the Research on Evaluation, Standards, and Student Testing

gdelacruz@cse.ucla.edu

Abstract. Educational videogames can be designed to provide instructional feedback responsive to specific actions. However, existing research indicates that students tend to ignore the feedback provided. That is, students often use ineffective help-seeking strategies. Research on the topic of help-seeking in learning environments have primarily focused on the role of cognitive factors, the nature of the help, or issues of timing and frequency. There is a noticeable gap in understanding how to motivate the use provided feedback. This study examined the relation between incentivizing the use of feedback and providing an explanation of the game's scoring rules on math learning in a pre-algebra videogame. A randomized-control design was used, comparing learning outcomes of students who received the incentive with those who did not. Results indicated that students given the incentive to use feedback had significantly higher normalized change scores on math items ($d = .53$), with stronger effects for students with low academic intrinsic motivation ($d = .88 - 1.17$).

1.0 INTRODUCTION

When event data are used to evaluate performance and have consequential meaning for the student, feedback is usually given. As such, one affordance of games for learning is that they can be designed to provide instructional feedback responsive to specific actions. Germane to the discussion of student use of feedback is the research on help seeking. The type of help that is often used in help-seeking studies is tailored hints and feedback, or supporting information accessible in general help menus. However, the research on help-seeking indicates that when students reach an impasse, they either use ineffective help-seeking strategies, or avoid seeking help altogether, as reviewed by [1].

To understand the relationship between help seeking and learning, researchers have primarily studied the role of cognitive factors (e.g., self-efficacy), the nature of the provided help (e.g., context-specific versus generalized principles), timing and frequency (e.g., on demand or system-initiated), and how the process of help seeking can be taught explicitly. There is a

noticeable gap in understanding how to *motivate* students to engage effectively in help seeking and use provided feedback.

Moreover, this study is focused on learning games where game play is directly integrated and linked with its corresponding academic content. If game play is to require the use, demonstration, and evaluation of the application of knowledge or skills in a particular domain, game students may need to know the criteria underlying their progress or lack thereof. Communication of a game's scoring rules can be leveraged to make assessment criteria more explicit and transparent.

This paper will describe a study that examined the impact of both incentivizing the access of feedback and providing different degrees of explanation of the game's scoring rules on math learning and performance in an educational videogame that was designed to teach students about fractions.

1.1 The Importance of Feedback for Learning

Feedback related to progress in learning tasks can generally have a positive effect on learning [2]. There are certain conditions that make it effective—specifically, feedback should support clear understanding of the learning goals or objectives (and the criteria that define good quality), a way for students' to relate own performance to that goal, and then a indicate the path to achieve the goal if there is any discrepancy between the goal state and their own performance [3]-[5].

However, the mere presence of feedback, hints, or available help is insufficient for learning. Review of the research on the use of feedback in technology-based environments indicates that students rarely access voluntarily available feedback. This is problematic as the lack of compliance with the treatment greatly reduces any potential effect on the outcomes. For example, Nelson [6] compared different levels of feedback on student achievement in an immersive learning environment. Results from the study indicate that most students did not access the feedback. Also, between students who were provided extensive feedback and students who were given moderate feedback, there were no statistical differences on the frequency of accessing the hints. In a study that examined the effect of providing user-initiated feedback via a pedagogical agent, Van Eck and Dempsey [7] also reported low levels of student access to the feedback. The authors concluded that further research is essential to figure out how to promote its use.

1.2 Increasing the Use of Feedback in Technology-Based Learning Environments via Incentives

The use of incentives for seeking help and feedback provides a relevant avenue for exploration. That students avoid help seeking is not surprising given that historically, it has been discouraged in most learning contexts. Research conducted in classroom settings suggests that there is social stigma attached to help seeking, a public sign of failure [8]-[10]. Further, in game-based settings, accessing feedback slows the game down, and some games have speed of play as a basis of advancement. Moreover, in some technology-based environments, the access of feedback or available hints are associated with lesser proficiency or may be discouraged. For example, when a hint is given to a student in the PACT Geometry tutor, the student's visible "skill bar" decreases [11]. In games, students are often penalized when accessing help. In is discouraged (i.e., students are penalized when a hint is accessed).

The design approach used in this study adopted the perspective that incentives can communicate what is valued within a given context and promote desired behavior. Rewards and incentives are, after all, a method for signaling those actions or behaviors that are encouraged or discouraged by a particular community. Conveying these expectations is a key process through which individuals learn how to participate in the communities that they are members of [12]-[13]. The use of an incentive to access feedback may be one potential approach, especially as findings from studies using tasks with initial low interest suggest that incentives may be beneficial [14]-[15]. Moreover, in video games, incentives are intrinsically tied to performance and are often a permanent aspect of the activity. In fact, game

designers argue that incentives (either in the form of rewards or punishment) are essential to the fun or sometimes the compulsion of the experience [16]-[18]. At best, an incentive can reverse the association between seeking help and failure, signaling to the student that seeking feedback is a valuable act leading to proficiency.

1.3 Rubrics of Scoring Rules to Promote Clarity of Expectations

In formal learning settings, scoring rubrics (or codified scoring rules and corresponding score values) have been used to improve assessment clarity by making explicit what constitutes good performance. There have been numerous studies on the teacher use of rubrics, typically for student-constructed responses. Most of the research has examined how best to train teachers to effectively use rubrics to increase the reliability and validity of scoring performance assessments [19]. Opportunities for student use of rubrics to improve learning appears logical, although only a few studies have examined this idea directly [20]-[22].

When the scoring rules are appropriately tied to academic progress, providing an explanation of the game's scoring rules functions as a rubric in a game for learning. The scoring rules make explicit the stated learning objectives of the game as well as the criteria used to evaluate performance. The scoring rules can direct attention to what and how responses are being scored which may (a) make more explicit the learning objectives or goals of the game, (b) clarify the criteria of performance (i.e., what "counts"), and (c) support the development of self-assessment of performance to determine when additional help is

necessary. This information can be provided prior to game play to guide performance, or as a context for elaborated feedback.

1.4 Summary

The review of the literature indicates for learning environments to be effective, students need to understand what constitutes good performance and use provided feedback in order to circumvent the ineffectiveness of trial-and-error learning. The impact of two aspects of design features in game-based environments were examined on math achievement, game play, and use of feedback: (a) incentivizing the use of feedback, and (b) testing of different levels of making performance criteria explicit through the explanation of scoring rules.

2.0 METHODOLOGY

The following section will describe the research question addressed by the study, the research design, descriptions of the treatment conditions and dependent measures used in the study and the analyses of the data.

2.1 Research Question

In a game designed to teach the addition of rational numbers, are there differential effects on math achievement measures, game performance measures, and access of feedback help based on the amount of the use of an incentive to seek feedback and the amount of explanation of the scoring rules provided?

2.2 Research Design

Data was collected from 112 students in fourth to sixth grades in after-school contexts. A randomized-control, 1×4

design was used in this study. There were four treatment conditions, which are described in the following section. Within participating after-school programs, each

Table 1 contains descriptions of the four treatment conditions used in the study with regard to the amount of scoring rules

Table 1. Description of Treatment Conditions Used in the Study

Treatment condition	Amount of information player was given regarding game's scoring rules				After a mistake was made, player was given an incentive to access additional feedback help
	<i>Explanation of scoring rules given prior to game play</i>	<i>Scoring explanations available in general help menu</i>	<i>Explicit feedback about when and how many points are earned and lost</i>	<i>Math elaborated feedback of scored event</i>	
No scoring information	O	O	O	O	O
Points-only feedback	O	O	X	O	O
Explanation of scoring rules	X	X	X	X	O
Rewarding help seeking	X	X	X	X	X

2.3.1 Experimental Materials: Save Patch

The game used in the study targeted two key mathematics ideas: (a) only identical units can be added to create a single numerical sum and (b) the size of a rational number is relative to how one whole unit is defined.

The objective of the game was to help the game character (Patch) jump over obstacles (e.g., spikes, lava, quicksand) and move from block to block to reach the last "X" block (the final goal). To do this, students

participant was randomly assigned to one of the four conditions.

2.3 Treatment Conditions

explanation that was provided and the incentive to use feedback.

needed to compute the distance of the jump, place trampolines on the blocks, and add enough coils to the trampolines to make Patch bounce. The size of the coil determined how far Patch would bounce. For example, a one-half unit coil would cause Patch to jump over a one-half unit interval.

The first part of the game required the student to determine the size of the intervals of the grid. The intersection of vertical red bars indicated the boundaries of the whole unit. The green dots broke up the whole unit into intervals (see Figure 1).



Figure 1. The intersection of the vertical bars depicted the boundaries of the whole unit. The green dots broke up the whole unit into intervals.

Once the students have figured out the unit size of the spaces on the grid, they needed to add together the correct number of coils that will span the distance to jump over. For example, in Figure 3, to get safely from one block to the next, Patch needs to jump over three one-third-unit intervals. This means that to successfully make it to the next block, the trampoline must contain three one-third-unit-sized coils. If the trampolines did not contain the correct number or size of

coils necessary to get Patch to the next block, Patch exploded into feathers and the students were allowed to replay the level. To reinforce the idea that only fractions with like denominators can be added together, the student could not combine coils that have different unit sizes. If a student had a trampoline with a one-third coil on it and tried to add a one-sixth coil, the one-sixth coil would not go onto the trampoline.

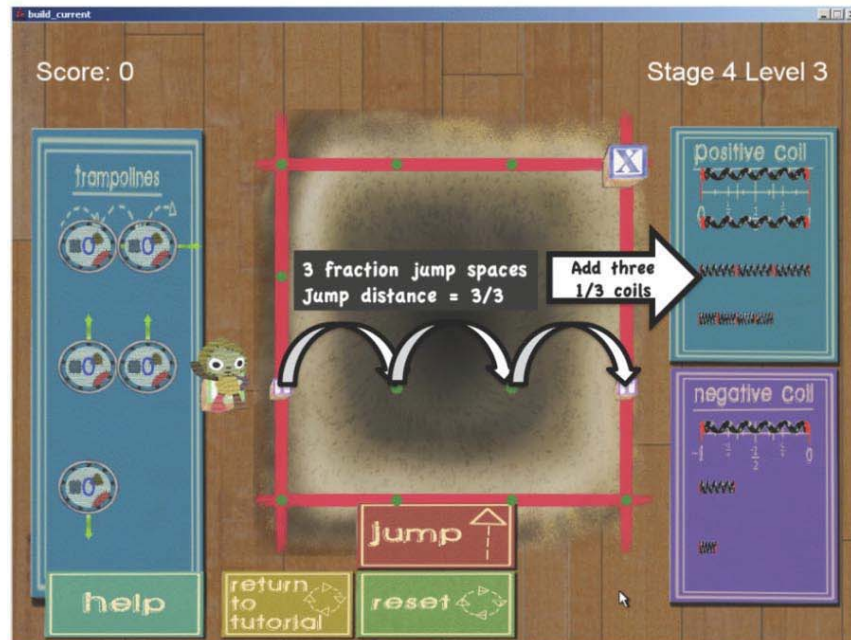


Figure 2. Jump distance equals total number of spaces between blocks. The number of coils to add to a trampoline equals the jump distance.

2.3.2 Save Patch Instructional Features

Tutorials: The design of the tutorial was meant to first contextualize the math concept within the game, and then discuss how that concept relates to math in general. The tutorial information was presented as both written text and guided interaction.

Scored events during game play: Across all of the conditions, three events were chosen as key points where performance would be evaluated because they mapped onto the learning objectives of the game. Points were *earned* any time the following event occurred: (a) student used coils that were the correct unit size for the grid, (b) student added together coils with like denominators, and (c) student successfully completed the level. Points were *lost* when any of the following events occurred: (a) student used coils that were not the correct unit size for the grid, (b) student attempted to add coils with unlike denominators together, and (c)

student failed to get Patch to an intermediate goal.

Feedback: All of the players in each condition received feedback after three events, that was in the form of the knowledge of results without elaboration or explanation. When a player tried to add a coil to a trampoline that had a coil with unlike denominators, the coils would not combine on the trampoline. Also, if the trampoline did not contain the correct-sized coil, or amount necessary to cover the distance of the jump, Patch exploded into feathers. When Patch jumped from block to block without exploding, this indicated that the player placed the correct quantity of coils needed to make the jump.

General help menu: All of the players in each condition had access to a general help menu at any time throughout the game levels. The topics that were included in the general help menu provided information on game mechanics (e.g., how to make Patch

move in different directions) as well as instructional information such as how to choose the right-sized coil and how to add coils of different sizes.

Feedback help. After the second consecutive mistake, all of the players in

each condition were given an opportunity to access additional feedback by clicking on a button that read, "Click here for help." The additional feedback provided elaborated explanations and hints that were designed to assist the player with repairing the mistake.

2.4 Dependent Measures

Table 3. Description of dependent measures used in the study.

Math achievement	Game performance	Total proportion of times additional feedback was accessed
Pretest: 31 items Posttest: 44 items (includes items in game context)	Number of: - coils added together - wrong-sized unit coils used - resets - failed attempts	- Overall in the game - After coils with different denominators were added together - When a wrong-sized unit coil was used - Failed attempt

2.5 Analyses

Three sets of orthogonal planned comparisons were conducted to examine if the data supported the hypotheses of the study.

3.0 RESULTS

Results indicated that compared to students who received minimal scoring information, students who were provided both explanation of the scoring rules and an incentive to seek additional feedback had significantly higher normalized gain scores on the math assessments ($d = .53$) and better game performance ($d = .99 - 1.09$), with stronger effects for students with low academic intrinsic motivation ($d = .88 - 1.17$). Students given the incentive accessed the feedback less frequently and spent the least amount of time on the feedback screens than the other students. However, when feedback was accessed, compared to the other students, they would solve the game level more quickly and with fewer mistakes. Furthermore, upon the

occurrence of the same mistake, unlike the students in the other condition, those given the incentive would seek additional information in the help menu on the same topic.

4.0 CONCLUSIONS

Overall, findings suggest that when designing games for learning, it is important both to make performance criteria (in this case, the scoring rules of a game) explicit to a student and to find ways to motivate students to use provided feedback through the use of incentives. There are two potential explanations of the positive effect of the combined incentive and explanation of scoring rules. First, providing the rationale for the scoring rules may clarify the criteria of performance, (i.e., what "counts"). Second, the incentive signals that the use of feedback is valuable, increasing the likelihood that students will engage in self-directed help-seeking and deeper engagement with the information on the screen. Findings from this study suggest that it is important both to make assessment

criteria explicit to students and to find ways to motivate students to use provided feedback through the use of incentives.

5.0 REFERENCES

- [1] Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277-320.
- [2] Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis* (Unpublished master's thesis). Vanderbilt University, Nashville, TN.
- [3] Bangert-Drowns, R. L., Kulik, C.-L., C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.
- [4] Hattie, J., & Temperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- [5] Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- [6] Nelson, B. C. (2007). Exploring the use of individual, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 83-97.
- [7] Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of skills in a computer-based instructional simulation game. *Educational Technology, Research, and Development*, 50(3), 23-41.
- [8] Karabenick, S. A., & Newman, R. S. (Eds.). (2006). *Help seeking in academic settings: Goals, groups and contexts*. Mahwah, NJ: Erlbaum.
- [9] Puustinen, M., Volckaert-Legrier, O., Coquin, D., & Benicot, J. (2009). An analysis of students' spontaneous computer-mediated help-seeking: A step toward the design of ecologically valid supporting tools. *Computers & Education*, 53(4), 1040-1047.
- [10] Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology*, 89, 329-341.
- [11] Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292-303). Berlin: Springer Verlag.
- [12] Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York, NY: Oxford University Press.
- [13] Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.

- [14] Cameron, J. (2001). Negative effects of reward on intrinsic motivation—A limited phenomenon: Comment on Deci, Koestner, and Ryan (2001). *Review of Educational Research*, 71(1), 29-42. Retrieved from http://www.unco.edu/cebs/psychology/kevinpugh/motivation_project/resources/cameron01.pdf
- [15] Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64(3), 363-423.
- [16] Fullerton, T. (2008). Game design workshop: A playcentric approach to creating innovative games (2nd ed.). New York, NY: Morgan Kaufmann.
- [17] Koster, R. (2005). *A theory of fun*. Scottsdale, AZ: Paraglyph Press.
- [18] Schell, J. (2005). *The art of game design: A book of lenses*. Burlington, MA: Morgan Kaufmann Publishers.
- [19] Baker, E. L., Abedi, J., Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research*, 89, 197-205.
- [20] Brown, G.T.L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system, *Assessing Writing*, 9, 105-121.
- [21] Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Researcher Review*, 2, 130-144.
- [22] Sadler, P. M. & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.

5.6 Why Games Work and the Science of Learning

Why Games Work and the Science of Learning

Curtiss Murphy
Alion Science and Technology
cmurphy@alionscience.com

Abstract. In 2010, the Navy formally added the Damage Control Trainer (DCT) to the recruit training program at Great Lakes, Illinois. Despite the incredibly dense training schedule at the Navy's boot camp, the instructors were willing to set aside two hours of time for recruits to play a game. Why? Because it worked. Even with just one hour of play, research showed that recruits gained a 50-80% improvement in performance that transferred to Battle Stations 21 (BS21), the Navy's capstone training event. This paper explores why games makes these kinds of results possible. It argues that the things that are known to improve learning are almost exactly the same reasons why games work: the time-honored laws of learning. It concludes that the traditional gulf between instructional design and game design is really an issue of perspective, rather than fundamentals.

1.0 INTRODUCTION

This paper will examine learning games, which is the use of games to introduce material, improve understanding, or increase retention. A lot of research has been done on learning games in recent years, and the results are sometimes nothing short of amazing. As an example, consider the Navy's Damage Control Trainer (DCT). In our studies of DCT, we showed a 50-80% improvement in individual recruit performance after just one hour of game time [1,2].

We performed a number of studies around the DCT that taught us a lot about building effective games, using games in a curriculum, and working in interdisciplinary teams [1,2,3,4]. That research has already been studied, collated, and published elsewhere. Therefore, this paper will not talk much about the DCT trainer itself or its development. Instead, it will present a new realization about game design that was uncovered while developing the DCT.

1.1 Thesis

Games like the Navy's DCT demonstrate that learning games can work. Therefore this paper will look past the question of 'can games work' and focus instead on the question of 'why they work.' It begins by exploring the laws of learning. These are the basic tenets that are known to improve learning outcomes [5,6]. Then, it will look at the basic tenets of game design - the techniques that lead to good games.

Comparing these two sets of tenets will uncover a surprisingly strong connection between learning and game design. In fact, it answers the original question. Games work because of the laws of learning. In other words, the things that are known to improve learning are almost exactly the reasons why games work. The goal of this paper is to help bring these two fields closer together by showing that the differences between learning and game design are mostly a matter of perspective, not of fundamentals.

2.0 THE LAWS OF LEARNING

Almost 100 years ago, Edward Thorndike described three basic laws of learning: readiness, exercise, and effect [5,6]. Since then, three additional laws have been added to the list: primacy, intensity, and recency [5,6]. Together, these six ideas are known as the laws of learning and they help us to understand how people learn and what conditions help them to learn better. These ideas have persisted through decades of research and study, and are still regarded as fundamental aspects of learning [5,6,7,8]. Even today, they are in use by instructors for the United States Army [8], Navy [7], Air Force [5], and Department of Transportation [6].

2.1 Law of Readiness

The law of readiness states that a student learns best when they are mentally, physically, and emotionally ready to learn. It says that motivation is a pivotal part of

effective learning. "Quite simply, motivated students learn more than unmotivated students." [7] This law involves student interest and the perceived value of the material. Further, it says that too many distractions can significantly reduce learning and that a student that is involved in the learning process will learn the material better. It also says that the best learning happens in conjunction with good physical health (rest, food, and exercise).

2.2 Law of Exercise

The law of exercise is not about physical exercise, as its name implies. Instead, it has two components that relate to exercising a particular skill. The first is that learning is increased through practice and repetition. The second is that feedback is important. Specifically, this law emphasizes that practice and feedback must exist together for the best learning results.

2.3 Law of Effect

The law of effect looks at the emotional responses of the student. It states that a student will learn more when the learning is associated with positive emotions. You can think of it as the law of positive feelings. As an example, a student who has success with initial material will associate positive emotions to the learning experience and will be more motivated in subsequent steps. Further, student choice has an impact on positive feelings and motivation.

2.4 Law of Intensity

The law of intensity states that things which are more intense are more likely to promote learning. This law is why we learn more from an exciting lecturer than we do from a book. It is why real life experiences make the best teachers – they are the most intense. In a nutshell, intensity heightens our perceptions and brings our full concentration to bear on a task, which increases learning.

2.5 Law of Primacy

The law of primacy states that the first thing you learn makes the strongest impression.

This law is why it is hard to replace flawed logic, unlearn bad habits, and fix negative training. This law is related to time.

2.6 Law of Recency

The law of recency says that learning degrades over time, so it is easier to remember the things we learned most recently. We can all remember material that was just discussed, but it is easy to forget a topic from last week or last month. This law is why instruction is designed to build upon prior material in a cyclical fashion and why we see significant declines in learning over inactive summer breaks [9]. This law is related to time.

2.7 Narrowing the Focus

These six principles are the laws of learning. When combined, they paint a picture of the basic conditions that are needed to improve learning. However, taken together, they cover an incredibly broad range of material, including everything from eating a good breakfast to the dynamism of the teacher to how often the student is physically exercising. While each of these laws is important, there is simply too much material to fit within the scope of this paper. Therefore, let's narrow the focus.

We will simplify this discussion by skipping over aspects that deal with the physical-world or with time-based conditions. After all, eating a good meal (physical) and what you learn first (time based) are important regardless of whether you are in a classroom or playing a game. Therefore, remove the laws of recency (time based) and primacy (time based) from consideration. Further, the law of readiness has physical aspects that can be eliminated, leaving only its motivational aspects.

This leaves only four laws and allows a closer look into some of the details. Specifically, let's independently examine the two components of the law of exercise: practice and feedback. Additionally, we will look at an idea that stretches across several of the laws: choice/involvement. We will call

these six focus areas the laws of learning for games (see Figure 1).



Figure 1 – Laws of Learning for Games

2.7.1 Motivation (from Readiness)

Motivation is a large part of the law of readiness and is strongly linked to learning in almost every way [5,10,11,12,13,14,15, 16]. Much research has focused on motivation and the results are very clear. A motivated student learns more, learns faster, and retains longer. Motivation is the holy grail of learning theory.

2.7.2 Feedback (from Exercise)

Feedback, like motivation, is strongly linked to learning [10,6] and is a crucial part of doing anything well [13,14]. Feedback can come in almost any form as long as it conveys your progression toward a goal.

2.7.3 Practice (from Exercise)

Time on task is a basic requirement of learning. Practice enables learning to occur and is a prerequisite of mastery. Research has shown that mastery requires 10,000 hours of practice [9]. It is important to note, as the National Research Council observes, "while time on task is necessary for learning, it is not sufficient." [10,p77]

2.7.4 Positive Feelings (aka Effect)

As expressed above, the law of effect can be thought of as the law of positive feelings. This paper will use that term. This idea simply states that learning is improved when associated with positive feelings/emotions.

2.7.5 Intensity

Intensity says that strong experiences will improve learning due to increased interest and heightened focus [7,8,5,17]. A dynamic speaker will hold your attention more than a dull one. Note that intensity is not limited to only positive experiences. Intense negative experiences can also be instructive, but are limited by the law of positive feelings.

2.7.6 Choice/Involvement (from Effect, Readiness, Intensity)

Learner choice and involvement are sub-parts of the laws of effect, readiness, and intensity [5,6]. Choice and involvement have long been shown to impact motivation [18], which is a part of readiness. From the law of intensity, we know that active involvement and choice are better than passive consumption. Finally, we know that being a part of the decision process that led up to the current material impacts both the feelings toward the learning (effect) and the motivation to learn (readiness) [18,11,12]. This sixth law puts all of this together into one idea in order to highlight its importance and enable comparison to elements of game design.

3.0 GAME DESIGN

Now that we have identified the six laws of learning for games, it is time to look at game design. Like learning, game design is a very large topic. There are hundreds of books and papers on the subject. In addition, game design is further complicated because the process of developing the game impacts the design [19]. A full exploration would require a discussion of many aspects of game development, including the art pipeline, level editors, working with programmers, developing in iterations, and play testing.

Fortunately, we're not interested in every aspect of game design. The goal of this paper is to build a bridge between game design and the laws of learning. Therefore, we can limit the discussion to the essential

aspects of game design that influence all games. This includes seven techniques: flow, feedback, simplicity, immersion and engagement, choice, practice, and fun.

3.1 FLOW

The first technique is **flow**. Flow is “the state in which people are so involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it.” [13] It was first described by researcher Mihaly Csikszentmihalyi [13,14], but has since been expanded upon by numerous researchers including Seligman [20], Schwartz [21], Deci [11], and Pink [12]. Flow is often described as the optimal human experience and is sometimes referred to by the term ‘engagement.’ [20,11,12] It can occur with any activity and has been linked to a person’s overall well-being [20].

Flow is widely accepted to be one of the fundamental reasons that people play games [13,17]. It is the essence of games. For game designers, the question is not whether flow is important, but, rather, how long you can keep your players in flow [22].

3.1.1 Requirements for Flow

There are four requirements for flow [13,14, 17]. Each is explained in relation to games.

- 1) Clear Tasks – Player understands what he is to do. Sometimes this means deriving his own goals.
- 2) Feedback – The game gives a lot of feedback about progression towards goals and whether the player’s choices are working or not. The feedback is usually immediate.
- 3) Balanced, attainable goal – The tasks should be somewhat challenging, but also achievable and not overly long.
- 4) Concentration – The player must be able to concentrate on the game and the game should avoid distracting the player away from the tasks. This can happen with complex interfaces,

intrusive story elements, or interruptions to gameplay.

3.1.2 Balance of Difficulty vs. Skill

Most of the requirements for flow are pretty straight forward. However, the third requirement is particularly complex and will require further explanation. In a nutshell, it says that a task must be simultaneously challenging and achievable. For games, this means that objectives must be challenging enough (and not overly long) or we become bored and lose interest. On the other hand, if the task is too hard, we become frustrated and anxious, and, once again, we lose interest and motivation. In addition, as we play the game, our skills will naturally improve and the challenges have to increase to match. When drawn out, this creates the **flow channel** (see Figure 2).

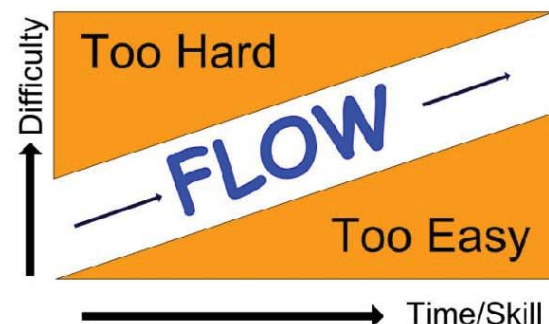


Figure 2 – Balance of Difficulty vs. Skill/Time

For games, balancing the flow channel is everything. It impacts motivation, learning, and enjoyment [18,13,14]. Since entertainment games are primarily for pleasure, a loss of player enjoyment or motivation is disastrous. Therefore, to keep players in flow as long as possible, the game industry has evolved an incredible array of techniques. Some of these include: chapter-based systems; dynamic difficulty adjustment; simplicity and transcendence; use of failure and repetition; eliminating the need for manuals; player adjustable difficulty; and expansive use of self-guided tutorials (both integrated and stand-alone) [3,17,23,24].

3.1.3 Link to Laws of Learning

By definition, flow is entirely about motivation [13,14], our first law of learning for games. Specifically, flow is about intrinsic motivation – the joy of doing [11,12]. When a player is in flow, he is completely focused on the task at hand, the tools and techniques available, and how he is progressing toward his goals. Given this level of intensity, it should come as no surprise that “the flow experience acts as a magnet for learning.” [14,p33]

3.2 Feedback

Feedback is the second technique used by game designers. It is also the second aspect of our laws of learning for games as well as the second precondition of flow. In short, feedback is important.

Simply stated, feedback is how you perceive progress [13,14]. It tells you when to stop pouring milk in a glass and how hard to push the brakes in a car. The glass fills up and the car slows. The feedback enables us to correlate our actions to outcomes. Without feedback, it is extremely difficult to perform even simple tasks (e.g. pouring milk into an opaque container with a lid).

The same is true in games. In fact, games are particularly adept at using feedback [17]. Fundamentally, games operate on a feedback loop [25]. A **feedback loop** involves four stages: 1) measure behavior; 2) relay the measurement to the user; 3) realize some sort of outcome; and 4) provide opportunities for alternate action.

The feedback loop has been an essential part of game design since the very first games gave points for eating ghosts/pellets. It's hard to even imagine what a game like Pac-man™ would look like without feedback. Imagine if the pellets did not disappear when touched, but you still had to touch all the pellets one time to finish a level. Feedback is such a core part of game design that modern businesses are now using the feedback seen in games as a way to improve real-life through ‘gamification.’

The use of feedback in games is fairly straight forward. But, at the same time, it can be applied in an almost infinite variety of ways. If you open any computer game ever invented, you might encounter typical feedback mechanisms such as: a scoring system; a way of tracking progress on goals (9 of 10 things); comparative statistics (high scores); an end-of-level debrief; growth indicators (50% through a cooking skill); and death/failure outcomes. Each of these gives feedback about your progression, performance, or skill growth. They show the outcomes your actions.

One final note about feedback in games is that it can generally be grouped into two categories: short-term and holistic. Short-term feedback is an immediate measure of how you are doing. For example, you hit something for 12 points of damage or you sold a gem worth \$50. Holistic feedback relates to the larger progression within a game. It has more meaning and weight. For example, maybe you unlocked a new dungeon area to match your increasing power or your character in the Sims™ got older. Holistic does not necessarily mean there is a delay between the action and feedback. It might still appear immediately, but it has more meaning than just +1 pellet.

3.2.1 Link to Laws of Learning

By definition, feedback is our second law of learning for games. In addition, feedback is a requirement for flow, which links it to motivation and makes it a critical consideration for game designers.

3.3 Simplicity

Simplicity is another core aspect of game design. Simplicity in games means that “everything should be made as simple as possible, but not simpler.” [27] The idea is not particularly new or even unique to games, but game designers consider it to be an essential, guiding philosophy. Consider this quote, attributed to Will Wright, designer of the Sims™, “Your garden is not complete until there's nothing else you can remove.” [28]

Games simplify the world down to “goals and rules of action” [14,p29] and allow players to focus entirely on “what should be done, and how.” [14,p29] This makes it easier for players to achieve flow, which increases motivation, and can lead to improved learning. For game designers, simplicity is not just an idea. It is a practical consideration that applies to all aspects of game design including the user interface, the goals of the game, feedback loops, game mechanics, user input, screen layout, story narrative, and the rules and instructions.

Another way to think about simplicity is the idea of **transcendence**. This means that “the player is more powerful in the game world than they are in the real world.” [17] At first, transcendence might seem to apply only to magical or fantastical settings. However, games, especially learning games, can be about anything in any setting. Even with mundane tasks and real world ideas, games offer transcendence. They simplify an experience and allow players to complete tasks faster than real life, with fewer obstacles, with less outside assistance, and in highly unlikely situations.

3.3.1 Link to Laws of Learning

By definition, simplifying means to reduce complexity. In games, this means reducing the number of steps, abstracting the process, limiting options, and reducing the cognitive load. By narrowing the set of goals and actions, a game makes goals easier to understand and helps players to correlate the feedback to their actions. In addition, games strive to present new material in its simplest form, which helps to balance difficulty versus skill. Finally, a simple user interface helps reduce distractions by minimizing extraneous information. All totaled, the simplicity offered by games has a direct impact on feedback and flow, which link to positive feelings and motivation.

3.4 Immersion and Engagement

The fourth technique comes from Douglas and Hargadon. They explained that there is

a difference between immersion and engagement [29]. **Immersion** is when you become deeply interested in a story or narrative. Immersion is usually a passive activity and can be associated with the idea of presence [30]. **Engagement**, on the other hand, is when you are actively involved in trying to work through a problem or puzzle.

As an example of immersion, consider watching a movie or cut-scene. As an example of engagement, consider playing a gripping online death-match or extinguishing a fire in the DCT. Though immersion and engagement overlap at times, the difference is that one is passive and the other is active. If you are immersed in a passive activity and it morphs into a problem you are actively trying to figure out, then you have become engaged instead.

The important point is that games often use both. They use a combination of story, attractive graphics, animations, particles, and other visual and audio techniques to immerse you passively in the virtual world [17]. At the same time, they engage your brain with a series of puzzles, choices, and problems that you must solve as part of the game experience. By combining these two, games can create very intense experiences.

3.4.1 Link to Laws of Learning

Not all games use immersion and engagement in the same way, but most of them use some combination of both. It leads to the powerful and compelling experiences that make games popular. In other words, games use the law of intensity.

3.5 Choice/Involvement

Games are the embodiment of **choice** and **involvement**. Sid Meier is often quoted as saying, “games are just a series of interesting and meaningful choices.” [31] Schell observes that not only must games provide choice, but the choices must be meaningful [17]. Not much else really needs to be said about how important player involvement and choice are to games.

Clearly, it is important, and clearly, games offer countless ways to do it.

However, there are other aspects of choice to explore. First, there is extensive evidence that choice and autonomy are associated with positive feelings and play a significant role in a person's overall health and well-being [11,12,21,20]. Second, despite all of the good that comes with choice, it has a dark side too: the paradox of choice [21].

The **paradox of choice** is that some choice is good, but too much choice can be very bad [21]. When faced with too many choices, people have difficulty weighing all of the factors involved. This leads to three possible outcomes: 1) we simplify the decision with increasingly arbitrary criteria; 2) we hit option paralysis (inability to decide); or 3) we postpone decisions. In games and in life, this is bad for several reasons:

- Too many options make a decision very complex, which increases the difficulty of the task. This impacts motivation and makes it harder to maintain flow.
- People are not good at comparing a lot of options. So, we simplify the criteria. In the extreme, the simplified criteria become almost random. This makes it hard to correlate actions to outcomes.
- We experience regret and buyer's remorse, which distracts us from the task and breaks flow.

Although this line of research is relatively new, game designers have long been aware of the need for simplicity (section 3.3) and thus, the need to limit decisions [17,24].

3.5.1 Link to Laws of Learning

The one idea of choice and involvement originates from three separate laws of learning: readiness, intensity, and effect. This idea means that people are more motivated, have more intense experiences, and have more positive feelings towards activities where they have choice, involvement, and control. By definition, a game is not a game without these things.

This ties game design to the sixth law of learning for games.

3.6 Practice

The sixth technique is **practice**. Games make extensive use of practice and **repetition** as a part of normal game play to help promote mastery [17,24]. Games generally require you to master a set of skills by overcoming a series of challenges that increase in difficulty and variety over time. New skills or techniques are added into the mix until you have mastered everything there is to learn and the game ends. As Koster points out, a good game is "one that teaches everything it has to offer before the player stops playing." [24,p46]

A common technique used in games is the idea of treadmills (aka. 'grinding') [32]. This is practicing the same basic sets of skills in order to level-up or improve her virtual character attributes. Usually, treadmills offer very slight changes in the environment or activity plus feedback such as experience or levels. Like a real treadmill, she is running in place with the tacit understanding that she is improving her skills.

In the Navy's DCT, repetition is used to strongly reinforce the importance of closing hatches once a Navy vessel is underway. The game teaches the player to close a variety of doors in a variety of situations. The lesson is practiced repeatedly and these skills transfer to the Battle Stations 21 graduation exercise [2].

Consider two final notes about repetition. The first is the possibility of emergent gameplay. This is when you have simple mechanics that combine in complex ways to create an incredible range of learning possibilities [24]. Emergent gameplay takes the idea of practice to a whole new level by opening up the ideas being explored within the game. The Sims™, League of Legends™, and Minecraft™ are great examples of emergent gameplay.

The second note about repetition has to do with the penny problem [33]. Games

sometimes use repetition in place of new content, as is seen in games like Tetris™ or Bejeweled™. In this case, you can end up with extensive exposure to material without real learning. The classic example is the inability to describe what's on a penny, despite having seen thousands of them [33]. This is an example where the techniques used in games may not improve learning.

3.6.1 Link to Laws of Learning

Games use practice and repetition in many of the same ways as traditional learning. Practice in games helps players achieve mastery of mechanics and helps players overcome increasingly difficult challenges in the flow channel. It is a direct application of the third law of learning for games.

3.7 Fun

The last technique is **fun**. It doesn't seem like a paper on game design would be complete without a discussion of fun. But what is 'fun'? It is pretty nebulous. It is used in lots of ways and means different things to different people, even to game designers.

Raph Koster says that 'fun is really just another word for learning.' [24,p46] He says that we play games because we enjoy finding, applying, and mastering patterns. Lazzaro, on the other hand, says there are four categories of fun: easy fun, hard fun, serious fun, and social fun [34]. Chris Crawford says that "fun is the emotional response to learning." [24,p229] Schell concludes that 'fun is pleasure with surprises.' [17,p37] These definitions have similarities but are each quite different.

However, there is one point all these designers agree upon: fun is an essential part of game design. It is a core consideration and a requirement for games. However you define it, regardless of how you achieve it, you must have it.

So, that still leaves us without a useful definition for fun as it relates to games. To solve this, we will define our own. Borrowing from the definitions of fun by Schell [17], Koster [24], Lazzaro [34], Crawford [24],

and Csikszentmihalyi [13,14,12], we can create the following definition:

Fun is the positive feelings that occur before, during, and after a compelling flow experience.

It is not perfect, but it is concrete. Plus, it links well to the ideas presented by Koster, Lazzaro, Schell, Crawford, and Csikszentmihalyi. The list of positive feelings associated with this definition of fun is quite long and includes: delight, engagement, enjoyment, cheer, pleasure, entertainment, satisfaction, happiness, fiero (triumph), control, and mastery of material. Best of all, it links the fun we experience in games to the idea of positive feelings associated with learning.

3.7.1 Link to Laws of Learning

Regardless of how you define it, fun links games directly to the law of positive feelings. In fact, most of the game design techniques described in this paper can be linked to positive emotions. Flow creates feelings such as engagement, satisfaction, and pleasure [13,14]. Immersion and engagement are both positive feelings, almost by definition [29]. Practice leads to mastery, which is linked not only to positive feelings, but also to overall health and well-being [20]. Choice and involvement are critical aspects of humanity and are an essential requirement for happiness [21].

4.0 TYING IT ALL TOGETHER

Section 2 introduced the laws of learning. These six laws describe the basic tenets that are known to improve learning. From that list, we eliminated the areas that were based on the physical conditions of the student (e.g., being healthy) and the time associated with the lesson (e.g., what is learned first). That left us with the six laws of learning for games.

Section 3 described seven critical techniques used by game designers. These include flow, feedback, simplicity, immersion

and engagement, choice/involvement, practice, and fun. Each of these aspects is important and ties game design to the laws

of learning. In Table 3, we combine and compare the two sets of ideas, side by side.

Law of Learning	Idea	Game Design Techniques
Motivation (Law of Readiness)	Motivated students learn more	<i>Flow</i> is the fundamental attraction of games. Games are <i>fun</i> and require moment-by-moment <i>choices</i> . This leads to extremely motivating experiences.
Feedback (Law of Exercise)	Feedback is how learners correlate actions with outcomes	<i>Feedback</i> is an essential part of games and a requirement for <i>flow</i> . The <i>simplicity</i> of games helps the learner correlate actions to outcomes.
Practice (Law of Exercise)	Practice is necessary for learning and mastery	Games use <i>practice</i> to promote mastery. They use increasing difficulty to keep players in <i>flow</i> and promote the learning of virtual or real skills needed to progress.
Positive Feelings (Law of Effect)	Learning is increased when associated with positive feelings	Games are supposed to be <i>fun</i> - defined as the positive feelings associated with compelling flow experiences. The <i>simplicity</i> and <i>involvement</i> of games encourages feelings of accomplishment and mastery.
Intensity (Law of Intensity)	Intense experiences increase learning, interest, and retention	A person in <i>flow</i> is intensely focused on an activity. The <i>feedback</i> loop intensifies the relationship between action and outcome. Games use a combination of <i>immersion</i> and <i>engagement</i> to create intense experiences.
Choice/Involvement (Laws of Intensity, Readiness, Effect)	Involvement and decision making can increase motivation, intensity, and positive feelings	Games <i>simplify</i> the world to a series of interesting and meaningful <i>decisions</i> . From moment to moment, players are actively engaged in the process of learning through experience.

Table 3 – ‘Laws of Learning For Games’ Linked to ‘Game Design’

Once the two sets are laid out like this, there can be no mistaking the strong synergy between them. In fact, one thing becomes extremely clear. Many of the principles that are critical for effective learning are also a fundamental part of games. Conversely, many of the techniques that make games appealing are also a part of what is needed to improve learning.

5.0 CONCLUSION

Certainly, this paper is not a comprehensive exploration of this topic. The science of learning is advancing quite rapidly and the laws of learning are just the beginning of what we know about how people learn. In addition, motivation and well-being have only recently become a major topic of focused research.

The same is also true of games. The art and science of game design is a complex subject that has only been briefly explored in this paper and, just as with learning, is rapidly evolving. The game industry is

exploding into a new world of social and mobile games and the gamification movement is attempting to blur the lines between games and our day-to-day lives.

There is much we still have to uncover about the science of both learning and game design. However, this paper has laid out some concrete, fundamental similarities that lead to the following conclusions.

First, game designers need to take a closer look at the science of learning. If a core part of games is learning, then game designers would do well to learn more about how people learn. If games work because of motivation, practice, and intensity, then designers should spend time learning what motivates us, why we practice, and the basics of human psychology.

Second, instructional designers need to take a closer look at games. If games can create significant improvement in performance, then learning designers would

do well to figure out why. If games work because of flow, simplicity, choice, engagement, and fun, then instructional designers should spend some time learning how to encourage flow, how to leverage simplicity, what makes an activity engaging, and how to foster fun.

The parallels are too strong for either discipline to ignore. Perhaps by studying games, we will advance the science of learning. And, perhaps by studying learning, we will learn to build better games. The one thing that is clear is that there needn't be a gap between the two professions. Game design and instructional design are fundamentally just two ways of looking at the same problem. Perhaps, the place to start is a bit of change in each of us – a change in perspective.

5.1 Acknowledgements

The author would like to thank the many organizations involved in the DCT project including Raytheon BBN, Alion Science and Technology, Intelligent Decision Systems, I.D.E.A.S., University of Central Florida – Institute for Simulation and Training, the Office of Naval Research, and most importantly, the dedicated sailors and instructors of the US Navy.

6.0 REFERENCES

- [1] T Hussain, C Murphy, et al. "Designing and Developing Effective Training Games for the US Navy". Published/Presented at I/ITSEC 2009.
- [2] C Murphy. "How to Build an Award Winning Training Game". Presented at I/ITSEC 2010.
- [3] C Murphy, D Chertoff, M Guerrero, K Moffitt. "Designing Effective Games – Flow, Motivation, and Fun". The Design of Learning Games, Springer-Verlag. (In press) 2011.
- [4] Hussain, Murphy, et al. "Development of Game-Based Training Systems: Lessons Learned in an Interdisciplinary Field in the Making". Published in Serious Game Design and Development, Premiere Reference Source, 2010.
- [5] "Train the Trainer Guide". The US Air Force, Training Systems Division. Authors unknown. 2004.
- [6] Aviation Instructor's Handbook. US Dept of Transportation, Federal Aviation Administration 2008.
- [7] "The Revolution in Training". The US Navy. Multiple authors. 2000.
- [8] "The Fundamentals of Instruction". The US Army. Author unknown. <http://www.usarmyaviation.com>.
- [9] M Gladwell. Outliers. 2008. Little, Brown, and Co.
- [10]. J Bransford, A Brown, R Cocking. How People Learn. National Research Council, published by the National Academy Press. 2000.
- [11] Edward Deci. Why We Do What We Do. Penguin Group, 1995.
- [12] D Pink. Drive – The Surprising Truth About What Motivates Us. Penguin Group, 2009.
- [13] Mihaly Csikszentmihalyi. Flow – The Psychology of Optimal Performance. Harper Perennial, 1990.
- [14] Mihaly Csikszentmihalyi. Finding Flow. Mastermind Series. 1997.
- [15] K Williams, C Williams. "Five key ingredients for improving student motivation". Research in Higher Education Journal. 2010.
- [16] Z Taheri. "An investigation of the effective factors on students motivational beliefs". Europe's Journal of Psychology 7, pp 62-80.
- [17] Jesse Schell. The Art of Game Design, a Book of Lenses. Morgan Kaufman, 2008.
- [18] Richard Ryan, Edward Deci. "Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well Being". 2000.
- [19] C Murphy. "A Low Cost Methodology for Achieving Joint Objectives – ONR Game". Published at I/ITSEC, 2005.
- [20] M Seligman. Flourish – a Visionary New Understanding of Happiness and Well-Being. Simon and Shuster, 2011.
- [21] B Schwartz. The Paradox of Choice. Harper Perennial, 2005.
- [22] J Chen. "Flow in Games". Master's Thesis, 2007.
- [23] J Beck, M Wade. Got Game – How the Gamer Generation is Reshaping Business Forever. Harvard Business School Press, 2004.
- [24] Raph Koster. A Theory of Fun for Game Design. Published by Paraglyph Press, 2005.
- [25] T Goetz. "Harnessing the Power of Feedback Loops". Wired Magazine, June 19, 2011.
- [27] Albert Einstein. Unknown source & date.
- [28] Will Wright. Quoted in multiple sources. Exact date and source unconfirmed.
- [29] J Douglas, A Hargadon. "The Pleasures of Immersion and Engagement; Schemas, Scripts, and the Fifth Business". 2001.
- [30] A McMahan. Immersion, Engagement, and Presence-a Method for Analyzing 3D Video Games (pp68-70). 2003
- [31] Sid Meier. Multiple sources including Koster¹¹
- [32] R Koster. "A Game Without Treadmills." 2007. Available at www.raphkoster.com.
- [33] D Hintzman, et. al. "Effects of Similarity and Repetition on Memory: Registration without Learning". Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995.
- [34] N Lazarro. "Why We Play Games: Four Keys to More Emotion Without Story". Proceedings of the Game Developer's Conference, 2004.

5.7 Virtual Worlds; Real Learning: Design Principles for Engaging Immersive Environments

Virtual Worlds; Real Learning: Design Principles for Engaging Immersive Environments

Dr. Sharon Wyly
Full Sail University
swyly@fullsail.com

Abstract: The EMDT master's program at Full Sail University embarked on a small project to use a virtual environment to teach graduate students. The property used for this project has evolved over several iterations and has yielded some basic design principles and pedagogy for virtual spaces. As a result, students are emerging from the program with a better grasp of future possibilities.

1.0 INTRODUCTION

The start of this project was simple. For students entering their first month of graduate work in the education program at Full Sail University, the goal was to introduce them to what a virtual environment was, get them to explore a little and see how they could use a virtual environment in their own teaching. The faculty along with the program director deemed it necessary to introduce students to a virtual environment. Hearrington (2010) pointed out the virtual environments were particularly important for teacher education programs as a way to help teachers meet the standards set in the International Society of Technology in Education (ISTE) published NETS standards for education. Specifically, Hearrington thought that virtual environments met the first two standards of inspiring creativity and using digital age experiences. This aligned with program goals, as well.

The students entering this program were approximately 80% public school teachers ranging from kindergarten through high school educators including public school technical staff. The remaining class was primarily higher education employees. Frequently, in an incoming class that ranged from fifteen to fifty, there were one or two corporate training specialists. All students entering the program were given a new Mac Book Pro to work on. Most of the students, particularly those who were public school teachers, were not sophisticated in their use

of technology; many had never used a Macintosh computer at all. Any use of a virtual environment needed to be easy to maneuver.

The virtual environment chosen was Second Life due to the fact that it had a large educational presence already in it, including ISTE and The New Media Consortium (NMC). Another important factor was that Second Life was stable on the Macintosh platform; not all virtual environments run on this platform. Also, as Foss (2009) noted, the global use and the strong economy provided a support that other virtual environments did not. Finally, Second Life had a simple download and installation procedure that was easy for students to follow, even for the ones who were not technologically accomplished.

Many institutions have a site in Second Life and the faculty members visited different campuses trying to find what worked in Second Life. Depressingly, many campuses seemed empty; it was rare to encounter another avatar. The most frequent design concept seemed to be to replicate the physical look of the campus, which was great for orienting people to what the physical campus would look like but did not seem to attract much use based on the empty spaces seen. Even the Full Sail campus was a replica of the real campus. The spaces encountered seemed cold and unappealing. Was there a better way to create a university campus in virtual space? Castronova (2007) referred to a *fungineer*

as a person who designed virtual spaces that were compelling and highly social using techniques pulled from psychology and game theory. This was the approach that the faculty was most interested in, and it lined up with the course within the program that taught game theory for education as well as other courses in learning theory and motivation.

2.0 ITERATIONS OF THE PROJECT

In June of 2009, the Educational Media and Design Technology (EMDT) began a new class in the master's program that was designed to be a front door course. The program is held totally online. As such, one of the desired outcomes was to introduce Second Life to the graduate students because it was used in several other courses throughout the program. Even though this was an accelerated course with month-long terms, it was decided that it was necessary to introduce students to a virtual environment early in the program.

In the third week of this new course students were asked to sign up for a Second Life account and to create an avatar. The students then met their instructor in an assembly in Second Life that was held in a public area. Since the class size ranged from fifteen to fifty students, it was a considerable crowd. Furthermore, since students created a Second Life account with whatever avatar name they preferred, the instructor had no clear idea of who was actually a student. This proved to be a difficult situation as anyone in the area could and did attempt to join in with the class trying to find out what was going on. The result was a disruptive environment that left both instructor and students unhappy. It was clear that the class needed a private place to gather. The following term, a more private place was sought but it still proved to be too public for a class. People outside of the class were attracted to the group and interrupted, in attempts to discover what was happening. The situation was not conducive to good pedagogical practices.

2.1 First Iteration

Due to the problems of being in a public space two months in a row, one faculty member made a decision to purchase a piece of land with the hopes that students from any course in the program would be permitted to use it. In order to own land, an account holder must have a premium account. Linden dollars, the Second Life currency, is exchanged on the world monetary exchange and the price in United States dollars varies. Prices reported here are in United States dollars and given the flexibility of the currency market, subject to vary. Most often, the exchange rate hovers around 200 Lindens per dollar. The lowest monthly charge for a premium account is \$14.95 per month and rises, based on the amount of land owned; however, the account holder also is granted an allowance of 200 Linden dollars per week, paid back into the in-world account.

The faculty member purchased an account and then paid for the initial investment in the land, house, landscaping and furniture. Total cost for all was under \$50.00. The original land purchased in Second Life was a fairly small parcel, about 1,024 square meters. The parcel was set high on a hill and, in the distance, the Linden Ocean could be seen. A small house was placed on the land surrounded by landscaping to create a pleasant environment. Included in the landscaping were a small pond and a butterfly garden. Inside the house, on the first floor, were a living room area and a dining room area. The upstairs level of the house was set up as an office. Some outdoor furniture was placed around the land and on an upstairs deck (Fig. 1).



Figure 1: First property – upstairs deck

Students in the first month of the program gathered here on one designated evening of the term. Signs were placed on the property to answer questions about the team assignment associated with Second Life. Students were to join with their teammates, read the signs and select a topic from the short list provided. The instructor then gave a representative from each group a note card with two locations on it that related to the chosen topic. Students then left as a team to begin exploring Second Life. Students were told that they could come back to the land at any time but records for the land showed very little involvement outside of this one class meeting.

2.2 Second Iteration

About three months later, after considerable thought, the land was redesigned. The house, which had only infrequently been used, was removed. A waterfall and stream were placed on the land so they flowed down to the pond with the only building a small gazebo on a bridge over the stream. This opened up the land and provided a more spacious look. It was easier for the instructor to visually track students without any walls impeding view but it was still not satisfactory. While the students expressed appreciation for the space, there was no reason for students to stay there, to talk, to be together. Also with twenty to fifty students on the land at one time, the place was crowded. It seemed to be time to move.

One factor in the decision to move was an expression by many of the students that

they wished they could go visit the ocean seen in the distance. Since the physical campus is located in Florida, although not near the ocean, students associated the university with Florida and Florida with the beach. Interestingly, this coincided with Theodore's (2009) exploration of Second Life where he found himself being drawn to areas that had water on them. So a search for an oceanfront property began.

2.3 Third Iteration

Early in 2010, a piece of property on an inlet close to the ocean was purchased. This new land was significantly larger (over 4,000 square meters) compared to the previous piece. Again, this purchase was funded privately, mainly with the account holder's allowance as the price for the property was well over \$100 just for the land. The property was also surrounded by other, very small pieces of land. In the ocean, were several small parcels that had tiny islands on them. This land carried a Linden Labs rating of G, which meant general use, and reduced the chance that neighboring lands would have any objectionable material on them.

Another change was made, this one to the assignment associated with Second Life. The initial assignment was for students to explore Second Life and report back on how a virtual environment might be useful to them. The outcome which was to make a multimedia presentation about Second Life was not in alignment with the purpose of the assignment which was to understand the how students could use a virtual environment to enrich their own teaching. Mayrath, Traphagan, Jarmon, Trivedi and Resta (2010) reported that relevance between the activity and the objective were a key factor in instructional design especially in virtual worlds. Students needed to see how virtual environments could enrich education environments. The assignment, therefore, changed to require students to conceptualize what a perfect learning environment for a selected lesson

would look like and to discover examples in Second Life that fit their view.

Mayrath et al also spoke to the time factor needed to be proficient in Second Life. Indeed, Falloon (2010) also found the same issue when he noted that the time needed to learn a new technology had to fit into the amount of time available. In an accelerated program, this was a challenge. The introduction to Second Life was moved from the third week of the program to the first week of the program to give the students more time to learn how to maneuver through the world. This meant that almost immediately, students were told that they needed to create a Second Life account and to begin thinking of educational spaces very different from the ones that they were currently teaching in.

The beach property was prepared for class use by placing a new house on the property, well back from the water. This house didn't have the rooms that the previous house had but rather was basically a one-story roofed platform. The entire house set up on stilts. Underneath the house, a small patio was placed and furnished with casual furniture (Fig. 2). This proved to be a favorite among students and was usually the first place students sat when the class started.



Figure 2: Casual sitting area

Down by the water, more casual furniture was placed in keeping with a general theme of a beach house. Tiki torches lit the area. This property also contained some space that was in the water, so the property was waterscaped as well with coral, plants and

fish. A small dock led out into the water and the landing point for the property was fixed there so as students arrived, they could see the entire property before them. Because of all this detail, the total budget for this move was close to \$200.

The property, which was used to host student meetings starting in July of 2010, was popular with students. During class, many students wandered around exploring the property and frequently used the beach area to stage scenes to be used in the class project. But once the initial class was over, few of them revisited the beach and subsequent Second Life assignments in more advanced classes did not generate any interest in going back to the property. Land records showed that most weeks, outside of the class meeting week, only two or three people visited the land. Some of these were, no doubt, neighbors who were wandering or who crossed over in their way to somewhere else. So, very few students were interested in coming back to use the property.

2.4 Fourth Iteration

In December of 2010, Linden Labs, the parent company of Second Life, closed the teen grid and moved those customers to the regular grid. At about the same time, Linden Labs changed the ratings on land to G (general use), M (moderate use) and A (adult use). At the same time, Linden Labs began to permit people as young as thirteen to sign up for accounts, but these accounts were restricted to G lands only. This helped to clarify the purpose of the property and to move adult sites off into areas of their own, which shielded younger users from stumbling onto questionable content.

This changed the way that educators interacted with the environment. Since previously some teachers had established places in the Teen Grid, they were now forced to move to the main grid or to move to different virtual environments. This actually made the decision to use Second

Life easier, since there was no longer a separation between the Teen Grid and the main areas. The faculty contemplated a move to a newer virtual environment but again, after a brief exploration, Second Life proved to be the choice because of the stability on the Macintosh platform and the simplicity of installing and starting. Another factor that weighed heavily in Second Life's favor was the in-world marketplace. Most of the faculty as well as the students were not able to build and script complex objects, so the marketplace allowed users to purchase or find for free any elements needed for projects such as chairs, trees or ducks.

At about this same time, changes needed to be made to the way the land was used. The beach property was a good start but the one drawback was that there was no room for multiple groups to meet at one time. Also, the land was getting full and it was difficult to add anything new. Since the property was surrounded by small plots of land that were for sale, it seemed to be a good time to expand (Fig. 3). The changes in Linden Lab's ratings actually meant that many people moved to the newly opened areas and that resulted in falling land prices. Some of the smaller pieces of land (512 square meters) were selling for only 1,400 Linden dollars (roughly \$7.00).



Figure 3: Land for sale

As pieces of land were put up for sale around the beach property, they were purchased and added into the main property. These smaller pieces of land were generally about 512 square meters, an ideal

size for a small group of four to six students to gather.

Gradually, land and ocean around the project came up for sale and was purchased and incorporated into the whole (Fig. 4) until the entire property contained more than 40,000 square meters. This increased the amount of items that could be placed on the land but raised land use fees for the account holder to \$195 dollars per month.

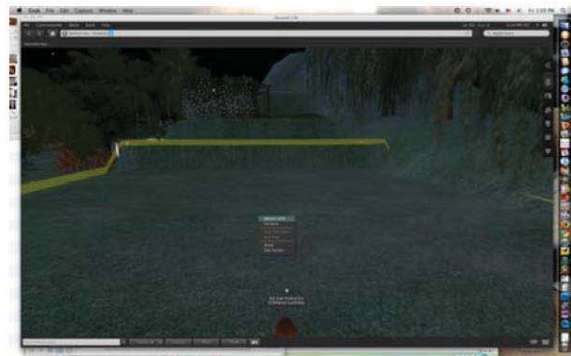


Figure 4: New land

The property was then re-broken up into smaller pieces of land that were used as collaboration spaces. This was necessary in order to create separate sound zones. This way, people could be working into two adjacent areas and not be able to hear each other.

In order to facilitate students finding their way around the property, small areas of land were given names and the landscaping and soundscaping was appropriately themed. For example, the Orchid Room (Fig. 5) had orchids all around the area, a small fountain and wicker furniture. Bird sounds were used in this area. Immediately adjacent was the Fern Room with a small waterfall and past that, the Crane Room had free moving crowned cranes and bird of paradise plants.



Figure 5: Orchid Room

Students could easily find an area to work in and the names also served for groups to use as a landmark to set up meetings.

As students progressed through the program, they began to use the land for multiple projects and for collaboration. Land use counts began to rise. It became more common to see weekly land use to be somewhere around twenty people.

Other faculty members within the program began to express an interest in using the Second Life land as well. Initially, one faculty member staked out a section as his virtual office and began to hold office hours there. More faculty members wanted to do the same but conflicts arose because students would anticipate that spaces would be open for their use only to find faculty members in there. A schedule for the land was rejected because there was a desire to keep the land as open and the use as fluid as possible.

As a result, it became even more important for the property to have some flexibility so that students had as many options as possible in order to explore and create.

2.5 Fifth Iteration

In order to have other faculty members fully participate in the Second Life experience, it was necessary to give them a space of their own. By this time, with all the land purchases, the property extended in a giant L shape and stretched down to a second beach area. The EMDT program director

owned a Second Life account for years and had been accumulating the allowance, so she donated 80,000 Linden dollars (about \$400) to the expansion. In March of 2011, a nine-unit apartment building was built on the second beach (Fig. 6) with a price of 10,000 Linden dollars (about \$50). The nine units each had two floors. The top unit was the penthouse and was substantially larger than the other units and included an outdoor pool and deck area.



Figure 6: Faculty Offices

This space was set-aside as a student collaboration area (Fig. 7). The decision was partially practical, to avoid having to make a decision as to which faculty member would get the penthouse but more importantly, as a symbolic reminder of how important the students were in this project. The penthouse was decorated with furniture and art; additionally, the property was landscaped to complete a modern tropical theme.



Figure 7: Penthouse

A lottery was held and, in numeric draw order, faculty members selected a unit. Initially, there were four faculty members. Each of the original four faculty members

received two thousand Linden dollars (about \$10) in order to purchase furniture. Three stores within Second Life were suggested, but the faculty members were free to buy wherever they wanted. This budget was enough for most of them to furnish their first floor but left the second floor of most units empty. Given all of the free items available in Second Life, many of the people were able to find things to fill up that space as well. Within a few months, though, two more faculty members signed on and another unit was set aside as a machinima studio. Students frequently has expressed interest in having a place to shoot like looked like a home environment. On the first floor of the unit, the machinima studio was furnished with a dining room set, a kitchen set and a living room set, in order to facilitate the creation of video projects within Second Life. The second floor of the unit was left empty so students could bring in whatever they wanted to make the place look like what they needed.

Not all of the faculty members liked the idea of having a space in the faculty building. Similar to the students, they expressed a desire to be outdoors. So, for one faculty member, Beth's Beach (Fig. 8) was built so she could have a space that reflected her personality. The area was designed to be an open yet defined office. One faculty member even contributed a photograph that was hung as art work on the wall.



Figure 8: Beth's Beach

Faculty members were free to use their space however they deemed. Some used it to hold virtual office hours; some explored

what a virtual environment was like and, still others used the space to create video materials (machinima) for their class.

With this final building, the land was essentially complete. Land use rose to about thirty to forty people per week, a considerable number for a program that has approximately 250 students enrolled.

3.0 DESIGN PRINCIPLES

Along the way in this project, several design principles evolved and were subsequently followed in designing new spaces and redesigning older spaces.

The first principle was openness. No walls were to be used whenever possible. Visually, the land was to be expansive as possible. The only building on the entire property with solid walls was the faculty office building.

The next principle was natural elements. People enjoyed being outside. Even when an interior area was provided, people gathered in outdoor sites. So as much as possible, any construction was avoided on the property and natural elements like plants or even fish tanks were placed inside whenever possible. Water was used throughout the property to provide a relaxing element.

The third principle was soundscaping. From the start of the project, sound was important. Linden lands come without any sound. In many ways, this was ideal because it allows for personalization. The first property had bird sounds and a slight wind noise. The move to the beachfront property meant adding in wave noises. Different collaboration rooms had slightly different soundscapes. In some places, elements such as a woodpecker that hammered on a tree have been added to enrich the environment.

The fourth principle was to encourage exploration. This was accomplished by

using small elements that were not easy to see with a quick glance. For example, down under the dock on the first beach, there was an octopus that moved around under the dock. Students had to wait for it to emerge or jump into the water and swim down to see it. This type of small detail meant that the longer one looked around the land, the more there was to see.

These principles clearly came into play when the designer of the Full Sail University Second Life campus entered the EMDT master's program. Kris Newton, a computer-programming professor at the university, was asked by the Marketing department to create a Second Life presence. His initial build replicated the campus (Fig. 9).



Figure 9: Full Sail campus – initial build

This was exactly what had been requested and Newton duplicated it right down to the rows of chairs in generic classrooms (Fig. 10).



Figure 10: Classroom

As part of completing his master's degree, Newton redesigned the campus using what he had learned throughout the program

about educational spaces and in particular educational spaces in Second Life. The redesign did not resemble the physical campus at all but reflected the spirit of the institution.

Newton's vision was lighter, brighter and seemed more inviting than the previous version (Fig. 11).



Figure 11: Redesigned campus

Furthermore, Newton created a classroom space that could swiftly be reconfigured at the touch of a button. The room could be arranged around a conference table or a desk and chairs or in a more casual open circle (Fig. 12).



Figure 12: Classroom in Conference Setting

4.0 DISCUSSION

As the project has progressed, several things have been learned. First, no one in the EMDT faculty had the expertise to

actually build or script objects. So everything had to be bought. The placement of elements on the property was a much bigger challenge than originally thought. Placing items in a virtual environment so they looked correct from all angles was not an easy task and took many hours of practice.

Another challenge was that the faculty members had various levels of Second Life expertise ranging from mastery to neophyte. Some had never been in Second Life at all. So a series of meetings were held to train everyone on the use of Second Life starting with how to create and dress an avatar and leading up to building objects. This grew within the university to become a series of professional development workshops open to all faculty and staff. Out of those workshops, a permanent round table was established, led by Newton, and open to all faculty and staff through the institution.

The land is always being reformed. Recently a waterfall was added and a new office established on a mountaintop for a faculty member. A request for a grotto underneath the waterfall is under consideration. Some of the students now use the property for their high school classes; other students come to the property just to relax. Recently, the project has been the site of several machinima projects. There is now a second piece of property at 20,000 square meters waiting for the right idea to come along and transform it.

At this time, the 40,000+ square meters between beach one and beach two are at maximum capacity. In order to grow, another move is needed. The options are to purchase a complete mainland sim or to purchase an individual island. Either of these options would yield about 65,000+ square meters. The move will be expensive and every option is being considered. When the next iteration comes, there needs to be a plan in place before the building starts. The current property was expanded bit by

bit and as a result, there has never been a master plan for where to put everything.

Another factor that was raised along the way was one of digital citizenship and ethics. A situation arose where students intruded onto neighboring land, even intruding into someone's Second Life house. When they were seen by a faculty member, their reasoning was that the property owner was there so they had not hurt anything by using this property. This led to many hours of discussion about what constituted fair use of virtual space. Finally, some of the faculty members drafted an acceptable use policy for the students using Second Life and this was shared among the whole program. It also opened up some interesting discussions with students about the issue.

5.0 CONCLUSIONS

This is clearly an example of a small project that grew much larger than the original creator ever thought it could. Starting out on a project like this does require some thought and some planning. The uses of the space need to be considered but the entire space needs to be kept as open as possible to create a dynamic environment that can be used in a variety of ways.

One thing that has been clear was that the vast majority of students enjoyed the Second Life experience. Many preferred to use Second Life to other means of collaboration such as iChat, AIM or Skype. Some students have used the land for relaxation or to unwind. Even after the initial class, students continued to use the property as a meeting spot for discussion and group work within their cohort. Others have brought their own high school classes into Second Life. Two graduates of the program created an immersive experience in Second Life to explain genocide to middle school students and bring their classes through that project. Additionally, one recent graduate reported that he had received and

accepted a job offer to work full-time in Second Life for an education company.

There are two potential pitfalls that can be made in Second Life and both were made by this project. The first is to attempt to define the space very rigidly. Allowing students and other faculty to give feedback along the way can create a property that is more open and flexible. The second is to try to replicate first life. Reality is not the way to go; with infinite possibilities, things that cannot be accomplished in reality should be tried.

The second life of the EMDT program has been a success. The next move will be to even more adventures in virtual environments. Students are passing through the program and going on to use Second Life and other virtual environments in a variety of ways, creating their own immersive environments for their students. The virtual is proving to be a reality in graduate education.

RESOURCES

Castronova, E. (2007) *Exodus to the virtual world: How on-line is changing reality*. New York: Palgrave MacMillan.

Falloon, G. (2010). Using avatars and virtual environments in learning: What do they

have to offer? *British Journal of Educational Technology*, 41(1), 108-22. doi: 10.1111/j.1467-8535.2009.00991.x

Foss, J. (2009). Lessons from learning in virtual environments. *British Journal of Technology*, 40(3) 556-560. doi:10.1111/j.1467-8535.2009.00955.x

Hearrington, D. (2010). Evaluation of learning efficiency and efficacy in a multi-user virtual environment. *Journal of Digital Learning in Teacher Education*, 27(2) 65-75. Retrieved from ERIC database.

Mayrath, M., Traphagan, T., Jarmon, L., Trivedi, A., & Resta, P. (2010). Teaching with virtual worlds: Factors to consider for instructional use of Second Life. *Journal of Educational Computing Research*, 43(4), 403-44. doi: 10.2190/EC.43.4.a

Theodore, P. (2009). Real education in virtual space: Looking at a multi-user virtual environment through a Deweyan lens. *Journal of Philosophy & History in Education* 59, 118-21. Retrieved from Education Research Complete database.

Acknowledgements: Thanks to Dr. Holly Ludgate, program director for EMDT for all her support, and to the entire faculty of the EMDT program for their unwavering enthusiasm. Special thanks to Kris Newton for his wonderful Full Sail campus work.

5.8 Metaverse Utilization in the EFL Classroom - A Study. Second Life® as Part of an Institution's Curriculum

Metaverse Utilization in the EFL Classroom - A Study Second Life® as Part of an Institution's Curriculum

Yukimi Imaoka, Yuki Amaki, Kenji Yagi, Chris Fuller, Anthony Madden
Kanda Institute of Foreign Languages

yukimi@kifl.ac.jp; amaki-y@kifl.ac.jp; heavyken@kifl.ac.jp; cifuller@kifl.ac.jp; madden-a@kifl.ac.jp

Abstract. Starting on April 2010, the Kanda Institute of Foreign Languages integrated the Metaverse application known as Second Life into the core curriculum for 840 new enrollments. As of April 2011, we enrolled an additional 760 students. This course, the Hybrid English Learning Program, is commonly referred to by the acronym: HELP. By taking the advantages of learning in immersive 3D environments, we aim to meet two objectives, 1) namely to enable students to remember and properly use learned phrases and expressions, and 2) using Second Life® as a tool to enhance the meaning and practicality of what is being learned. Although there have been several online English learning programs or services here in Japan, this endeavor represents the first time Metaverse technology has been utilized as part of a core curriculum; a curriculum structured upon classroom learning and interaction within the Metaverse. This paper endeavors to explain the following: the planning and development of HELP, the class structure, and the results of a questionnaire which measured whether using such an approach would reduce student timidity and assist student to gain more confidence in speaking English in different social environments.

1.0 INTRODUCTION

Using metaverse applications in the learning of a foreign language has three prime benefits. First, the learner can create their own avatar to represent themselves, but at the same time shroud their true identity, allowing for anonymous interaction. Therefore those who are shy or introverted can be more active and effective speakers.

Second, role-play can easily be done allowing for many varied situations. The situations they encounter and the speaking practice they acquire in the metaverse can easily translate into real life situations. This is very important since actual conversation is not static, but dynamic, not always following a set pattern.

Third, using the metaverse, students can venture outside the walls of the classroom, study remotely, or even study along with students from other countries. In addition, this can help students acquire the skills needed to effectively communicate using other internet-related services.

Based on the benefits mentioned above, Starting on April 2010, the Kanda Institute of Foreign Languages (KIFL) integrated the

metaverse known as Second Life® into the core curriculum for all new students. This course, the Hybrid English Learning Program, is commonly referred to by the acronym: HELP. At the start of the 2010 Academic year, new enrollments at KIFL totaled 840. As of April 2011, we enrolled an additional 760 students.

In this paper, we will talk about the structure of HELP and we will evaluate the results of surveys we conducted among the students. This survey primarily measured the almost psychological effect of how using an avatar alleviated the timidity that many students experience when speaking a foreign language.

2.0 PLANNING AND DEVELOPMENT

By taking advantage of the metaverse for use as a language learning tool, we aim to meet two main objectives, 1) namely to enable students to remember and properly use learned phrases and expressions, and 2) using Second Life as a tool to enhance the meaning and practicality of what is being learned.

Language functions primarily as a tool which accompanies us in our daily life; useful in the numerous situations we find ourselves in.

Logically, we cannot expect that students will gain practical effectiveness in speaking a foreign language if they only study the target language. Hence it is common in EFL, English as a foreign language, classrooms for students to do role-play activities, whereby they practice using the knowledge they have acquired. Undoubtedly, these activities are beneficial. However, role-play activities of this nature have some limitations.

For example, while practicing a role-play conversation, the students must imagine the setting they are acting out, which rarely captures the real sensation of the actual situation they may encounter. Moreover, the very nature of practicing with another EFL student can sometimes result in learning a rigid conversation pattern, stifling the development of natural conversational skill.

Therefore, with these limitations in mind, we have combined the classroom lesson with the metaverse. By using set dialogs and other important skill building exercises practiced in the classroom, along with metaverse activities, we created a hybrid conversation-based curriculum.

Learning English in a non-English speaking region can make it difficult for the student to perceive the practical value in their classroom lessons. By means of this integration of the traditional and the virtual, we believe we have created an environment more conducive to the overall improvement of the EFL student's practical English ability. Furthermore, when students converse and utilize the expressions and dialogs they studied in class in a virtual reality setting that copies its real-world counterpart; our hope is that they will more easily make the connection between the hypothetical and the actual.

As they see how using English helps them to accomplish various activities in the virtual world, we hope they will experience the

satisfaction of finally 'getting it,' seeing the value of learning English, feeling the joy that comes with seeing their own advancement, and as a result, develop the much sought after and essential trait necessary for successful language utilization: self-confidence. We therefore hope that this hybrid method will become an alternative way for foreign language acquisition.

3.0 CLASS STRUCTURE

The entire course curriculum is an integration of three essential components; each component being the fusion of two elements: 1) the utilization of the classroom and the virtual online world of Second Life, 2) the collaboration and class oversight by Japanese and native English instructors, and 3) the combination of traditional learning styles and amusing activities. Below is an outline of the actual course procedure.

(1) In the classroom, Japanese instructors teach and explain to the students the English conversation dialogs (phrases, expressions, grammar structure, improvisation techniques) and cultural background essentials.

(2) In the metaverse setting, native English teachers converse with students via their avatars. This in-game experience mimics the real world and persuades the students to use what they have learned in order to progress through the given scenario.

(3) Review and discussion of student's Second Life in-game experiences.

4.0 ANONYMITY

4.1 Shyness

For students who are studying a foreign language (for example: English) in a non-native environment (such as Japan), the key element that they need to master their studies is having a setting where they can practice their speaking and listening skills; helping them to build adequate fluency, memorization, and improvisation. Therefore in the typical EFL classroom, role-play activities are

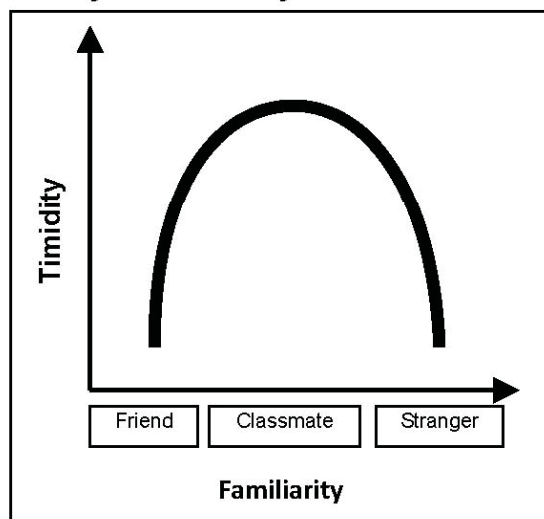
essential. Students stand in front of the classroom and practice the target dialog or situation with another student.

However, many students feel a bit shy in doing such role-play before the class. In addition, since they are speaking with a fellow Japanese student they feel somewhat or even highly embarrassed to speak English with them. This situation can sometimes hinder the effectiveness of said role-play activities among Japanese students.

4.2 When Does Timidity Manifest Itself?

In 2005, Reference [1] the book, *Shuuchishin ha Doko he Kieta? (Where Does Timidity Go?)*, written by Sugawara contained a study concerning shyness in Japanese society. It said that amongst the Japanese timidity is based on one's familiarity with the person they are interacting with. The study showed that with the Japanese, interestingly, they do not feel particularly shy when interacting with close friends or family, nor with complete strangers.

Timidity and Familiarity



However, the Japanese as a whole feel extremely timid when interacting with people they know, but who are not their friends, for example, their classmates, colleagues, or neighbors. When interacting with such individuals many Japanese hesitate to

express themselves or engage in more than just simple small talk.

4.3 ANONYMITY

As mentioned, since the Japanese feel shy in certain social circles, it is understandable why many desire anonymity when using internet communication services.

Facebook®, the world's most popular SNS, as of May 27, 2011 boasted having some 3,480,000 accounts in Japan. However Facebook is not the most popular SNS in Japan, mixi® is. In Japan, as of April 14, 2010 mixi was reported to have some 20,000,000 user accounts.

There is a big difference between these two competing social networking sites. Facebook strongly suggests that its users create accounts using their real names. However, mixi allows its users to create accounts using only nicknames. Therefore, due to the anonymity that mixi allows, many Japanese feel more comfortable using this SNS rather than Facebook.

Reference [2], the survey done by Mobile Marketing Data Labo was reported on April 6th, 2011.

In the below table, people were asked how they identify themselves when using social media services. Only 23 percent of the users of mixi had made their real names public. In contrast, on Facebook, some 84.3 percent of its users make their real names public on that social network.

The three major social media websites (mixi, Twitter, Facebook)

The level which users share real information

	Facebook	Twitter	mixi
sharing a real name, school, or company name	49.0%	9.4%	11.1%
sharing only a real name	35.3%	13.6%	11.9%
using a nickname that your friend would know	7.1%	43.1%	56.2%
using a name	4.9%	27.9%	18.1%

that no one would know for anonymity purposes			
other	3.7%	6.0%	2.7%

By looking at both of these surveys objectively we can confidently say that when using online services, anonymity is very important to the Japanese as it helps them to communicate better.

4.4 Avatar as a Physiological Mask

By taking advantage of the "anonymity" or the detachment from self, a mindset commonly experienced while using avatars in a virtual setting, we hope to reduce the level of anxiety and embarrassment peculiar to the Japanese when speaking in a foreign language.

5.0 QUESTIONNAIRES

5.1 Questionnaire Overview

The purpose of this research is to provide insights into the Hybrid English Learning Program's effectiveness in using Second Life by revealing the positive and negative aspects of it from the perspective of current students. In the survey, we included questions regarding each student's country of citizenship, English language proficiency level, and overseas experience.

In addition, questions measured student experience in English communication in real life as well as in Second Life. The questions focused on whether using an avatar to converse in English helped them to overcome shyness and embarrassment or not. Our hope was that by their responses we could accurately determine that using said avatars in Second Life would help English learners overcome cultural barriers to speak English with more confidence.

5.1 The Questionnaire Participants

The target population for this study was our current first-year and second-year students

enrolled in HELP. The total number of those able to take the survey was 785.

5.3 Questionnaire Procedure

On KIFL INFOWEB, our vocation school's internet website which student access to receive general school information and other important information peculiar to HELP, a questionnaire form was created for use by the HELP students. The participating students were asked to fill out the questionnaire as honestly as possible.

5.5 Questionnaire Period

From July 6, 2011 to July 12, 2011

5.6 Data Analysis

The 785 students who responded to this survey came from more than six countries. The majority of respondents, 95 percent, were Japanese citizens. 2 percent came from China and 1 percent from South Korea. 476 first-year students, 61 percent, and 309 second-year students, 39 percent, responded to this survey. 44 percent of respondents had spent less than one month abroad, and 39percent of respondents had never been overseas.

After identifying themselves in this way, the next questions assessed ways to overcome the barrier of shyness when speaking the English language in real life and in a virtual reality environment. When we analyzed 785 students' responses using agree-disagree questioning methods, there were significant differences in overcoming the language barrier of shyness in speaking English through avatars and in speaking English in real life (see Table 1-1).

The survey included twelve questions for overcoming shyness in speaking English. There were no significant differences in speaking their native language with strangers classmates, teachers, or friends in real life or in Second Life. The data shows that most students do not feel shy to speak their native languages.

According to the survey results, See Table 1-1), 46 percent of students feel shy to speak in English with strangers, and 53 percent do not feel shy to speak with strangers in real life. Through an avatar, 69 percent of students indicate that they "somewhat disagree" and "totally disagree" with being shy about communicating with strangers in Second Life.

The virtual environment is one significant way to overcome their language barriers of shyness. Also, 66 percent of students "somewhat disagree" and "totally disagree" with being shy about communicating with teachers in the real-life setting. Through an avatar, 71 percent of students "somewhat disagree" and "totally disagree" with being shy about communicating with teachers in virtual reality. Thus, the virtual-reality environment has played an important role in aiding students to overcome the language barrier of shyness in speaking the English language.

Another important finding is that fear of embarrassment is one of the negative factors in overcoming the barrier of shyness. 60 percent of students indicate that they "totally agree" and "somewhat agree" that they fear embarrassment if they make a mistake while speaking English with strangers in the real-life setting. Through avatars, 59 percent of students do not feel embarrassed when they make a mistake while speaking English with strangers in virtual reality. On the other hand, 58 percent of students indicate that they "somewhat disagree" and "totally disagree" that they fear embarrassment when speaking English with teachers in real life. Through avatars, 68 percent of students do not feel embarrassed if they speak English with teachers in virtual reality. The data indicates that through avatars, most students do not feel embarrassed when speaking English with classmates as well as friends. Communication in English in the virtual setting may help students overcome the language barrier of shyness and the fear of embarrassment.

Table1-1
Shyness in Real Life versus Shyness in Second Life

	Conversation in the Real Life		Conversation in Second Life	
	"Totally agree" and "Somewhat agree"	"Somewhat disagree" and "Totally disagree"	"Totally agree" and "Somewhat agree"	"Somewhat disagree" and "Totally disagree"
1. Do you feel shy to speak your native language with strangers?	23%	77%	13%	87%
2. Do you feel shy to speak your native language with teachers?	16%	84%	11%	89%
3. Do you feel shy to speak your native language with classmates?	12%	88%	12%	88%
4. Do you feel shy to speak your native language with friends?	9%	91%	10%	90%
5. Do you feel shy to speak English with strangers?	46%	54%	32%	68%
6. Do you feel shy to speak English with teachers?	28%	72%	29%	71%
7. Do you feel shy to speak English with classmates?	35%	65%	22%	78%
8. Do you feel shy to speak English with friends?	25%	75%	21%	79%
9. Do you feel embarrassed if you make a mistake while speaking English with strangers?	60%	40%	41%	59%
10. Do you feel embarrassed if you make a mistake while speaking English with teachers?	42%	58%	32%	68%

11. Do you feel embarrassed if you make a mistake while speaking English with classmates?	41%	59%	37%	63%
12. Do you feel embarrassed if you make a mistake while speaking English with friends?	36%	64%	29%	71%

6.0 CONCLUSION

According to the data, student survey responses indicate that in many cases communication in English through avatars leads students to overcome the language barrier of timidity and the fear of embarrassment when speaking English with strangers, teachers, classmates, and friends. This positive effect on the students leads us to the conclusion that metaverse applications and the use of avatars in the EFL classroom makes for an improved language learning curriculum.

7.0 APPENDICES

HELP Online Questionnaire Survey

Conversation in Real Life

1. Do you sometimes feel shy when speaking your native language with the following people?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				
With classmates, etc				
With friends				

2. Do you feel shy to speak English with the following people?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				

With classmates, etc				
With friends				

3. When you make a mistake while speaking English, do you feel embarrassed?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				
With classmates, etc				
With friends				

Conversation through Avatars (Virtual Reality/Second Life®)

4. When communicating through an avatar, do you feel shy to speak your native language with the following people?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				
With classmates, etc				
With friends				

5. When communicating through an avatar, do you feel shy to speak English with the following people?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				
With classmates, etc				
With friends				

6. When communicating through an avatar, do you feel embarrassed if you make a mistake while speaking English with the following people?

	Totally agree	Some what agree	Somewh at disagree	Totally disagree
With strangers				
With teachers				
With classmates, etc				
With friends				

REFERENCES

[1] Sugawara, Kensuke, *Shuuchishin ha Doko he Kieta?*, Kobunsha Publication, 2005.

[2] *The three major social media websites (mixi, Twitter, Facebook) The level which users share real information*, MOBILE MARKETING DATA LABO, 2011.
http://mmd.update.ne.jp/news/detail.php?news_id=784
(Accessed on July 15, 2011).

5.9 Crossing the Virtual World Barrier with OpenAvatar

Crossing the Virtual World Barrier with OpenAvatar

Bruce Joy
VastPark
bruce@vastpark.com

Lori Kavle
VastPark
Lori.Kavle@vastpark.com

Ian Tan
VastPark
ian.tan@vastpark.com

Abstract

There are multiple standards and formats for 3D models in virtual environments. The problem is that there is no open source platform for generating models out of discrete parts; this results in the process of having to “reinvent the wheel” when new games, virtual worlds and simulations want to enable their users to create their own avatars or easily customize in-world objects. OpenAvatar is designed to provide a framework to allow artists and programmers to create reusable assets which can be used by end users to generate vast numbers of complete models that are unique and functional. OpenAvatar serves as a framework which facilitates the modularization of 3D models allowing parts to be interchanged within a set of logical constraints.

1.0 INTRODUCTION

Virtual world, game and sim developers often want to be able to generate Avatars (models which represent users in a virtual world). Until now there has been no Open-Source code project to generate 3D models with sufficient quality. Unless sufficient money was spent to create a proprietary system for that virtual world, users were limited to pre-generated models. This has been the case in many Enterprise virtual worlds and military simulations. This added the burden of creating all the end user assets on the 3D artists and led to multiple end users sharing the same avatar model due to the limited choices available and not having a photo realistic avatar to match their face. There is interest in the user's association with the avatar regarding immersion [1,]. Thus avatars with the user's



Figure 1. Collection of Female Avatars

face tends to increase the immersion in the environment for all the users.

Modularity is a major benefit in the process of generating models. For example, a human model could have endless combinations of garments that the model could be representing. This is not just a simple matter of changing the skin, but also changing the mesh of the model; simply swapping out the skin for a man wearing a suit to jump suit will be detrimental to the aesthetic of the model and its believability. There are strong demands and advantages to having models where the components can be swapped out according the context. Take an instance where a user may have their face modelled to their avatar. If they switch contexts, such as moving from being a pilot to a doctor, they would still be able to keep their face. This would give the user a greater sense of ownership and allow them to associate more with their avatar, further immersing them within the virtual world. OpenAvatar is not just limited to Avatars which are representations of humans in a virtual environment; they can be used to model anything from furniture to vehicles to animals.



Figure 2. Avatars from leading provider

There is a need for a high quality 3D model generator that uses an industry standard which can be accepted by tools commonly

used by artists. There are other vendors who offer high-quality models and face-capture technologies; however there was no end-to-end solution which incorporated these technologies in a reusable way.

Another problem was that models were not easily customised in-world, which limits the reusability of the models. There existed no open source means of separating parts of a model from the whole and swapping them in and out. This limits the reusability of models and creates more work for artists developing assets for 3D environments.

Currently there exists no open source software for compiling models from their discrete parts; OpenAvatar plans to fill this niche.

3.0 SOLUTION

OpenAvatar was designed to solve the issues mentioned earlier. It provides an open specification which allows artists to separate models into discrete components which are then bound to a skeleton. OpenAvatar uses COLLADA (.DAE) [2] files which is an open format for 3D models. COLLADA files can be easily exported by popular tools used in industry such as Maya and 3DS Max.

The process starts by having the artist create a whole 3D model and binding it to a skeleton for animation, though not all models will be animated, the skeleton is used as the foundation on how the model is to be assembled by the OpenAvatar compiler.

Collada is used due to its open specification; more importantly the XML structure of COLLADA allows it to be stitched together by the OpenAvatar compiler. Also, it can easily be converted

into the runtime format specific to that environment.

As not all 3D models are static and will be animated; Collada supports the BVH animation format conforming to standards used in the industry.



Figure 3. Comparison of OpenAvatar and Leading provider

4.0 OPEN SOURCE

Releasing the OpenAvatar SDK publicly under an open source MIT licence is integral to its adoption. The source code is made available for modification and integration with existing platforms. As there are a wide variety of platforms and use cases it make sense to have the code released so that developers and tailor it to their applications.

The OpenAvatar public SDK is entirely free to use for commercial and non-commercial purposes alike. We hope this free model serves as an incentive for adoption and testing allowing the public SDK to develop and suit the needs of users and developers alike.



Figure 4. Collection of Male Avatars

5.0 PROCESS

The process for creating a model for use in OpenAvatar is very similar to creating a model for any other 3D interactive platform. However for the model to be compliant to OpenAvatar it must have certain design considerations

The Model must have a skeleton; this is a given if the model is animated, however even if the model is static it will require a skeleton to be compatible with OpenAvatar.

The artist starts by creating the model as they normally would in the tool of their choice. After they create the model they bind the discrete components to the skeleton.

The skeleton is what guides the OpenAvatar compiler. It makes sure that the right parts go in the right places. The skeleton is important; however it does not work alone. OpenAvatar also uses XML files to dictate the compilation of the model.

The first of these XML files is the Archetype file. Archetype files serve as the contract between OpenAvatar and the artist. Archetype files stipulate what a model must have; they also stipulate what is optional for

that model. For example a model of a human must have a head, however whether that model has a backpack would be optional.

Archetype files are contracts, they may seem constrictive, but they actually allow for more variety and re-usability with regards to 3D models. For example a 3D platform could have an Archetype for a human, with a standard set of animations. Users can then upload their custom models which comply with the archetype file. Similarly mature, developed archetypes scan be used as a specification to artists who may be subcontracted, giving them a firm direction on how their models must be designed. The benefits from this are many. A client may not be restricted to one provider for their models, as long an artist can familiarise themselves with the archetype they can create the model. Artists can also create a model based on the archetype confident that it will integrate into the environment.

The next XML is a DNA file. The DNA file is the actual implementation of the Archetype file. The DNA file contains the references of the DAE files and which part of the model those DAE files belong to. For example a discrete part such as the left leg will refer to the leg DAE file. DNA files also contain the references to the skins of a particular component.

When compiling the model, OpenAvatar checks the DNA and the Archetype files against each other. The compilation will happen only if the required parts of the model are included and there are no parts included that should not be there.

6.0 BENEFITS

Re-usability is a tenet of OpenAvatar. To have faces of individuals swapped out of a

model would save time and money. This serves another qualitative benefit. User avatars which can retain their faces serve to further immerse the user in the virtual world or simulation.

Encapsulation is another aspect of OpenAvatar. Protection of intellectual property (IP) is a consideration when it comes to 3D models. If client wishes to have accessories for a specific model made; however does not want to release IP to the designer, it can be problematic. Archetype files can rectify this by providing the framework for which the 3D artist can create that accessory.

Though Collada is used in this server-side application, there is no need to release Collada files to end users as many real time 3D engines use custom runtime mesh formats. These formats can usually be generated automatically from the Collada file. This presents an additional opportunity where OpenAvatar generates Collada models that are then generated into runtime models for multiple platforms.

The resulting Collada file may be stored by

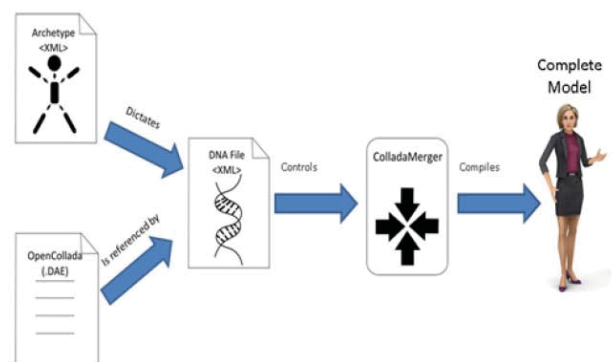


Figure 5. Overview of Compilation

the system to be used for future regeneration of runtime files or as an archive if desired.

Finally, OpenAvatar abstracts the end user decisions away from any mesh format. In the event that procedurally generated models become possible, then the OpenAvatar DNA files could be used to regenerate user avatars utilizing the new procedural approach. Users can seamlessly experience a content quality upgrade without having to go back and remake their decisions.

7.0 WHAT'S NEXT

OpenAvatar lends itself to many possible implementations. The DNA and Archetype constructs allow for standardization which can streamline collaborative work generating models.

A virtual world could have a web interface which would allow users to upload their custom avatars and models to give them greater creative control over how their world looks and feels. This would give a greater feeling of ownership to the user.

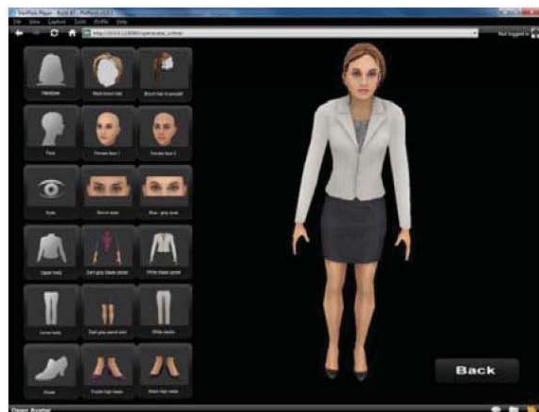


Figure 6. Interface for customisable avatar

While not every user of a virtual world is a 3D artist, OpenAvatar could serve an interface where models can be taken apart, have their parts replaced and reassembled.

The benefits to this are that avatars can be specifically assembled by users who do not come from a 3D artist background.



Figure 7. Facial expressions in action

The DNA file serves to facilitate re-usability; specific components can be hosted on the internet. An example use case could be that people's faces are hosted on a server whereby later they can be downloaded and modelled onto a different outfit. The benefits are that time and money can be saved as well as giving the user greater association to their avatar.

It is possible for the Archetype file to facilitate a marketplace. The Archetype file enables a concrete definition of component parts in an open specification; with an open specification and the use of Collada, 3D artists can create discrete components to popular archetypes and sell them. Conversely those looking to buy can put out a tender for a specific part to a model can release a "mannequin" which will serve as a base for 3D artists to create parts for the model. The open specification will allow any 3D artist to submit their design to the client for their consideration. For example Company X is looking for some new

backpacks for their models; they put up a mannequin which serves as the base for which this new backpack will be designed. Artists can then submit various models for the client's approval. The use of Collada and Archetypes would allow for "Crowd-sourcing" of 3D models. Note, the marketplace is enabled by OpenAvatar specifications but the SDK itself does not provide a marketplace.

8.0 CROSS-PLATFORM COMPATABILITY

Using Collada as the format for storing models in OpenAvatar, it inherently supports the ability to transfer an avatar to another virtual world. As a result user's can move between virtual worlds while retaining their avatar. By supporting this, it could create a greater exchange between virtual worlds; users will now have greater ease to retain their avatar and identity. They would not have to "start afresh" if they wanted to visit someone else in another world to share an experience. Collada files can be used to import models into second life.

A variation on that concept is the ability to retain their face; technologies like FaceGen allow a user to give their avatar their face. Though COLLADA can be converted into the run-time format of the platform, transferring a whole model from one environment's art direction to another might not make sense. For example, taking a user's avatar from a fantasy environment where he is a barbarian warrior over to more modern contemporary environment where he may be playing the role of a policeman would disrupt the believability of the world.

OpenAvatar's ability to separate a model into discrete parts allows a specific part to be taken and replaced onto the model which suits that world. The Archetype and DNA

files can facilitate this and allow it to be done procedurally and automatically. This means that the user need not configure his avatar to the new world, or some controlled configuration on the user's part can be allowed; to extend our previous example, the barbarian warrior may require a haircut before he can don the policeman's uniform.

9.0 CONCLUSION

OpenAvatar is an open source software and specification that aims to streamline the development and integration of models into virtual worlds. By using COLLADA which is an open specification, models compiled by OpenAvatar can easily be converted into the run-time format of the virtual world they will be deployed in.



Separation of models into discrete

Figure 8. An OpenAvatar avatar inworld

components will allow more reusability of 3D assets. Instead of building whole models, individual parts for the model can be built and be swapped in and out as the situation dictates. Similarly a model has a greater ease of being customised by the user or the virtual world.

Archetype and DNA files serve as contracts to standardise models ensuring that they are built to a certain specification, but also give security, ensuring that models will adhere to the skeleton and then can be rigged and animated when deployed in the virtual world. An agreed upon archetype can

serve as a standard for which additional parts and accessories can be created by 3D artists, enabling a wide variety of parts and accessories for that model. Using COLLADA as the format for storing the models, OpenAvatar aims to be compatible with various virtual world providers. With portable avatars the ease at which users can transfer between worlds increases, allowing for greater interaction between virtual worlds.

To conclude, OpenAvatar hopes to enable greater ease of generating models procedurally and give a specification with the Archetype files. Doing this will give the virtual worlds industry a framework with which to collaborate and grow as an industry

10.0 REFERENCES

1. Salomon, Mandy. (December 2010), Smart Services CRC, Immersive Internet Australia: Education, http://www.smartservicescrc.com.au/PDF/Immersive%20Internet%20Australia_Education.pdf, Page 27
2. COLLADA's website, www.collada.org

11.0 ACKNOWLEDGEMENTS

The Authors wishes to thank the closed beta volunteers who have tested out the OpenAvatar SDK and the following people who were instrumental in creating OpenAvatar as an open useful format:

Jeremy Massey, Lead, VastPotato (SG)

Craig Presti, CTO, VastPark

Chantu Shamsu, Developer, VastPark

Alec Voang, 3D artist, VastPark

And the rest of the VastPark team

5.10 Visual Tool for Authoring Learning in Games (VITAL)

Visual Tool for Authoring Learning in Games (VITAL)

Talib Hussain, John Ostwald, Kerry Moffitt

Raytheon BBN Technologies

thussain@bbn.com, jostwald@bbn.com, kmoffitt@bbn.com

Abstract. While the field of serious games has made great progress in creating effective products and in identifying some of the design methods that lead to success, there are few tools to help the developer of a serious game for learning. Most instructional logic, such as performance tracking, guidance, and feedback, is typically developed directly within the core programming language of the game engine or its scripting language. The Visual Tool for Authoring Learning (VITAL) is an open-source tool for creating the instructional logic for a learning game. VITAL provides the capability for instructional system designers and developers to collaborate via high-level interfaces, explicitly specify instructional elements such as learning objectives and instructional methods, and generate functional instructional logic without the need for low-level programming. It has been designed to be platform-independent and has been integrated with multiple game engines. This paper will introduce the motivations behind the tool's design, describe how VITAL interfaces with the game to provide an effective learning experience, and walk through some examples of using the tool to create instructional logic.

1.0 INTRODUCTION

Creating an effective serious game for learning (or “learning game”) typically requires significant design and development effort due to the intrinsic multi-disciplinary nature of the effort. Not only must one create all the usual game elements, such as game mechanics, art and sound assets, animations and interesting story and characters with appropriate player engagement and flow, but one must also create meaningful instructional elements, such as guidance, performance assessment, feedback and scaffolding with appropriate instructional logic, sequencing, context and learner motivation. Over the years, developers in the entertainment games industry have created a rich set of tools to help facilitate the game development process – from game engines [1-4] that encapsulate an array of graphics, physics and user interface mechanisms, to scene-level editors and game-specific modding tools [5-8], and 3D modeling and animation tools [9-11]. Additionally, game developers often use generic software development tools such as integrated development environments [12-13], scripting languages [14-16] and defect tracking and code analysis tools [17-18].

But, what is available to help facilitate the development of learning games? Essentially, the answer to date has been “Whatever tools are available to entertainment game developers.” Thus, to create all the instructional elements of a game, the

designers and developers currently must directly program those elements within the game using general programming and scripting languages. As a result, instructional logic and elements are typically “buried” deep in the game code and are not easily retrieved when analyzing the system, nor readily open to modification when enhancing the system. Some learning game efforts create instructional infrastructures to support robust instruction within their game, but these are generally ad-hoc in nature and are not usually re-used in subsequent efforts nor made generally available to other developers.

The Visual Tool for Authoring Learning (VITAL) was developed to fill this gap and provide a development tool to facilitate directly the creation of instructional content within learning games. VITAL was developed with the explicit intention of blending instructional constructs and concepts typically used by an instructional system designer with the constructs and concepts used by game developers. Its key goals are the following:

- (1) Support designers and developers from both instructional and gaming disciplines by providing them a collaborative authoring environment.
- (2) Introduce a level of rigor into the development of instructional content in games by embodying best practices.
- (3) Make instructional content a “first-class citizen” in the software, including natural support for embedded assessment.

- (4) Reduce the level of effort needed to create effective instructional content in games.
- (5) Directly support the re-use and straightforward revision of instructional content.
- (6) Do all the above in a manner that is agnostic to specific gaming engines, independent of any specific domain of instruction, and supportive of a broad range of instructional methods.

VITAL has been developed over a period of three years, is a fully functional authoring tool that has been used in developing multiple games on multiple gaming engines, and is available as open-source code at <http://sourceforge.net/projects/vital>.

This paper introduces the authoring approach behind VITAL, describes the infrastructure that supports explicit control of instructional logic within a game, and introduces the specific authoring interfaces and mechanisms used within the tool. We provide examples of using VITAL to create instructional content, and close with a discussion of the extensibility of VITAL.

2.0 VITAL OVERVIEW

VITAL is based upon the authoring concept that there are elements of the learning game that “belong” to the gameplay which should be developed using the game engine tools and should be controlled by the game engine – but that there are also elements that “belong” to the instruction which should be authored using an instruction authoring tool and should be controlled by an instruction engine. This approach extends the typical componentized approach of a game engine. Within a game engine, for instance, a physics engine controls physical interactions, a graphics engine renders graphics, and, perhaps, an AI engine controls character behavior. VITAL adds an instruction engine component.

Underlying our approach are several key observations:

- (1) A learning game involves elements that are solely related to instruction, elements that are solely related to gaming, and elements that involve both.
- (2) Within a typical learning game, interactions with the player – both gaming and instructional – are largely driven by the actions of the player.
- (3) As a player’s experience in a learning game unfolds, there may be shifts, as perceived by the player, from gaming elements to instructional elements. However, a well-designed learning game is first and foremost about learning. As such, upon inspection, it should be clear how gaming elements are being applied to serve the learning objectives.

Consider a rough analogy. Imagine that you have an instructor hovering over your shoulder watching as you play a game. As you play the game, the instructor tweaks things in the game and whispers helpful things in your ear. There are times during the game when you as a player may be performing activities that are below the threshold of interest of the instructor and other times when the instructor is deeply involved. For instance, in a game that teaches (among other things) how to bargain while shopping in a village, the instructor may pay careful attention to your activities once you enter a shop in the game, but have no interest in your activities when walking between shops and choosing what to buy next. While talking to the shopkeeper, the instructor makes sure that the shopkeeper gives you some challenging choices, and if you make too many poor bargaining decisions, the instructor whispers some feedback in your ear. But, the instructor may not be concerned with what the shopkeeper looks like, what the shopkeeper is wearing, or about a myriad of other aspects of the visual and interactive experience.

Imagine now that you also have a “game master” watching over your other shoulder. As you are walking between the shops, the game master makes sure that there are other characters walking around to make the area

more interesting, plays some background noises that are realistic of a small village, and even makes it hard for you to walk quickly if you have purchased a lot of things since you are burdened down. Once you enter the shop, the game master focuses on making sure that the shopkeeper looks, moves and behaves in a realistic fashion, but cedes the primary responsibility for the content of the shopkeeper's interaction with the player to the instructor. As soon as you complete your purchase, the game-master takes a keen interest again and makes sure that you can pick up and put the item in your inventory (perhaps even with a cool animation) and leave to walk around the village again.

Rather than creating all the logic behind both the "instructor" and "game-master" in the core game logic, VITAL provides the capability to specify the "instructor" separately from the "game-master," and to explicitly specify the interplay and handoffs between them.

Of course, this analogy is not perfect; as noted, some elements concern both the game and the instruction – and some instruction may come through game elements not entirely under VITAL's control. But even for such elements (e.g., if the instructor in our example really did care about the appearance of the shopkeeper) VITAL is useful. In cases where it is not providing instruction to the game directly during execution, VITAL establishes a framework that provides guidance to the development of those game elements it does not directly control (like the physical appearance of characters).

2.1 Infrastructure

Figure 1 illustrates the VITAL infrastructure. VITAL is developed in the Java programming language and contains an Instruction Engine that communicates with the game engine during execution (which may operate either as an independent process or embedded within the game process). The game engine sends events to the Instruction Engine, and the Instruction Engine in turn sends commands to the game engine. The events are typically directly based on user actions or

changes in game state. The commands are typically requests to the game to initiate or terminate certain interactions or change certain game states. The two engines communicate via an execution bridge that appropriately translates messages to/from a form understood by the game engine from/to the form understood by VITAL.

VITAL also includes an Instruction Editor, which allows the learning game author to produce a specification of all the instructional logic to be executed by the Instruction Engine. This logic is converted to a scenario data file in eXtensible Markup Language (XML) format, as opposed to compiled code. The XML file is interpreted by the Instruction Engine. This is analogous to the manner in which many scene level editors produce a data file that is interpreted by the game engine.

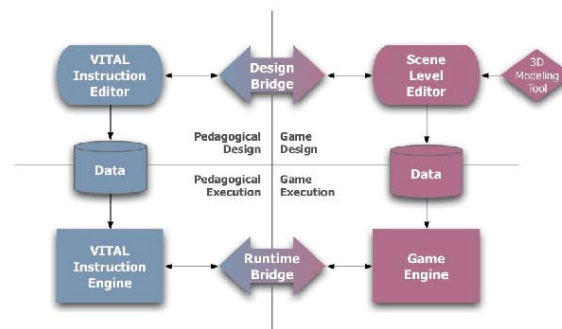


Figure 1. VITAL Infrastructure

2.2 High-Level Editing

In order to ensure that the tool is more easily accessible for an instructional system designer and to promote collaboration, the Instruction Editor focuses on specifying instructional logic at a high level via visual interfaces such as wizards, pull-down menus and drag-and-drop actions. Several different high-level interfaces are available within VITAL, and the core ones are arranged as tabbed panes. Figure 2 illustrates one of the editing tabs of the VITAL Instruction Editor that allows the author to explicitly associate learning objectives and narrative with specific instructional logic and sequencing. The details of the editor's capabilities will be discussed in later sections.

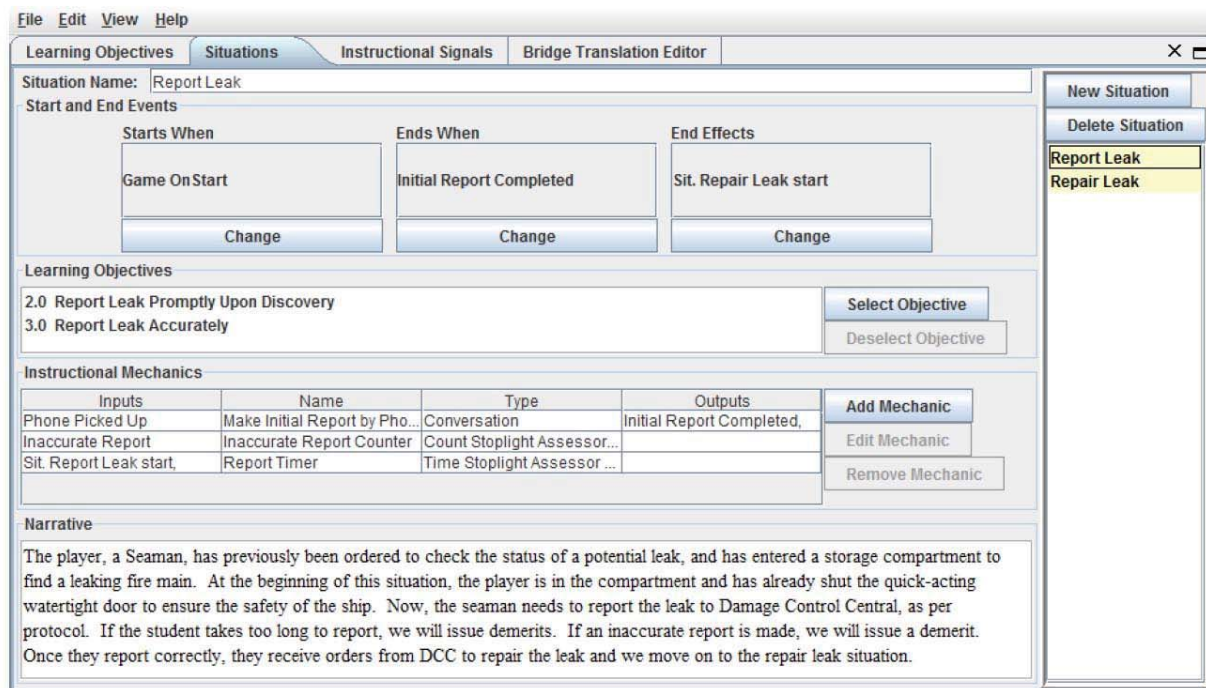


Figure 2. VITAL Instruction Editor's high-level, tabbed interface for specifying instructional elements

2.3 Visual Logic Language

At a lower conceptual level, close to the game engine, the Instruction Editor is grounded in a visual logic language that we developed specifically to use interaction-driven logic and instructional logic components. Within this language, the instructional logic for a game is depicted as a graph of nodes and connections. The nodes represent events, instructional components and logic operations (e.g., and/or gates). The connections among the nodes represent the flow of action and information during execution, and, indeed, developers can debug their logic at this level, watching state changes and signal flow in real-time in a graph running "live."

Through the appropriate configuration of nodes, the language can specify arbitrary conditional branching and/or repetition logic. For instance, Figure 3 illustrates a sample graph in the language. The graph shows that, when the user performs a specific action (e.g., opening a door to a

room on fire), it will trigger one of two outcomes. If the player's character is wearing appropriate safety gear, then a positive reward will be given. If not, then a warning feedback will be issued.

While the visual logic language supports the creation and debugging of highly complex instructional logic directly within a visual editing interface, the true strength of VITAL is that all authoring actions performed in the high-level interfaces are automatically compiled into the appropriate visual logic graph elements. Ultimately, the graph logic is exported as an XML scenario data file to be interpreted by the Instruction Engine.

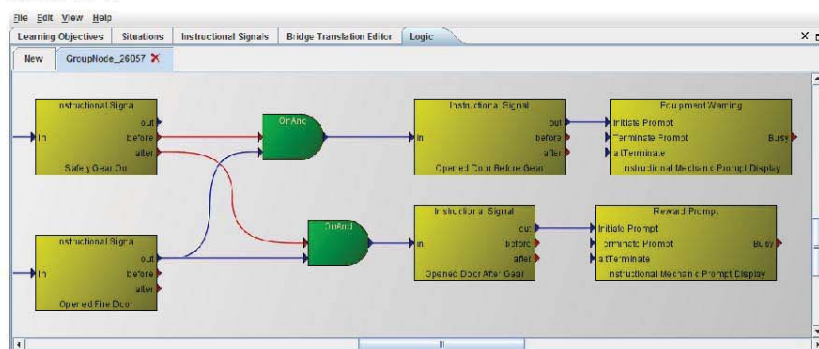


Figure 3. Sample graph in VITAL's visual logic language

2.4 Communication with Game

For a given VITAL scenario, the Instruction Engine communicates in real time with the game engine in which that scenario is realized. This communication occurs through a set of signals to and from the game engine, relayed across a bridge from the Instruction Engine to the game engine. The bridge is comprised of two portions – a data-driven part that may be modified using interfaces in the Instruction Editor, and a hard-coded part that is specific to the game engine.

Within the data-driven portion, the authors first define a domain ontology using whatever terminology they prefer. A domain ontology is simply a hierarchy of nouns (types of things, such as *door*) and associated verbs (types of actions per type, such as *open*) that allows for as much description as the instruction needs to know about what is happening in the world of the game.

Next, the authors establish a set of instances comprising specific instances of particular nouns. These correspond to the actual objects within the game world about which the instruction cares to know, and/or whose behavior it will manipulate.

If necessary, the authors may define a separate ontology and instance set for a particular game engine, and then use the Instruction Editor to define translation links between the domain ontology and the game object ontology. This allows, for instance, the same domain ontology (and all higher level logic using that ontology) to be used with different game engines that have different game object ontologies.

Within the game engine, support must be developed to translate the kinds of events that the game engine knows about into the signals that VITAL understands, and vice versa. This hard-coded bridge development is necessary only once per game engine for a given game object ontology.

This approach to communicating with the game was adopted in order to support domain independence as well as platform-independence of VITAL.

3.0 AUTHORING APPROACH

Now that we've described the basics of how VITAL works, let's dive down into exactly how VITAL allows a user to author instruction. In order to achieve our fairly ambitious goals, VITAL uses several novel concepts and mechanisms that blend traditional instructional system design thinking and typical game design thinking. Throughout, however, an overriding aspect of the VITAL authoring approach is the explicit specification of instructional elements, and a respect for best design practices of both stripes.

3.1 Learning Objectives

As any good instructional system designer will tell you – it all begins with the learning objectives. In designing a learning game, it is essential to understand what is being taught and what level and type of performance is expected of the learner. VITAL respects this priority by requiring the author to explicitly specify the learning objectives for the game. Figure 4 illustrates the learning objectives editing interface. Objectives are entered directly as plain text by the author. The learning objectives are numbered and arranged hierarchically. This allows the author to quickly and easily enter in the objectives they have identified during the instructional system design process and set the foundation for the game's instruction.

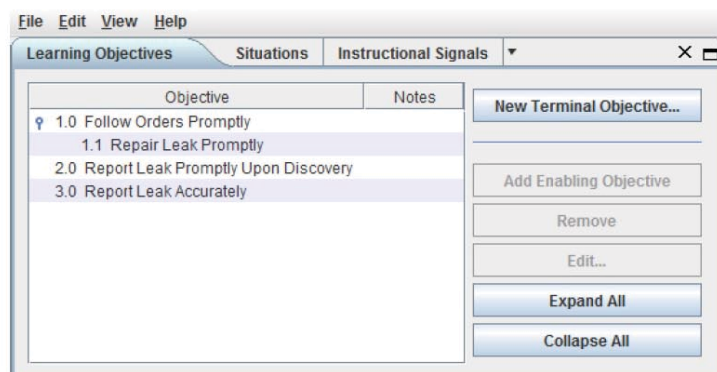


Figure 4. Learning objectives editing interface

3.2 Situations

While having a set of learning objectives is a good start, it is just that – a start. A key aspect to an instructional design is the order in which those learning objectives will be taught. This is termed instructional sequencing, and a variety of approaches may be used – such as teaching the parts of a task before teaching the whole task, or teaching simpler concepts before introducing more complex ones.

In a game, in an analogous manner, a good game designer will often break up the gameplay experience into smaller experiences. These may focus on different aspects of the evolving game storyline (e.g., first gain experience, then fight the dragon, and finally save the princess), or different elements of a complex task (e.g., to solve the mystery, you need to collect clues and solve smaller puzzles first). They may also break up the gameplay into a progression of increasingly complex tasks (e.g., the traditional concept of “levels” of a game).

Best practice in the design of learning games suggests that a good approach will blend these methods to break gameplay up into a set of experiences that build upon each other by focusing on different learning objectives and different parts of the gameplay.

Within VITAL, “Situations” play this role. A Situation captures a particular experience in the game within which a specific learning objective or set of learning objectives is taught using a specific instructional method or set of methods. VITAL allows the author to specify a set of situations and to order them appropriately to reflect a desired sequencing of instruction and gameplay.

Figure 2 above illustrates the interface for defining a Situation. The Situation is associated with specific learning objectives – namely *Report Leak Promptly Upon Discovery* and *Report leak Accurately*. The Situation begins when a certain condition is met (namely the game has started). During the Situation, the instruction is controlled by a set of instructional mechanics that collectively control specific instructional interactions

(such as a conversation) as well as embedded assessment (e.g., tracking how long they take to report) and feedback (e.g., providing feedback when they make errors in their report). Instructional mechanics are defined in more detail in a later section. The Situation ends when the player achieves a certain condition (namely, they have completed their initial report of the leak).

An important aspect to note about Situations is that they reflect a set of logic that is executed in parallel with the game and are effectively overlaid on the game experience. Going back to our earlier analogy, we can imagine a Situation as reflecting a single instructor focused upon a specific set of learning objectives looking over your shoulder. Moreover, multiple Situations may be active at the same time, analogous to multiple instructors looking over your shoulder, each focused on teaching you something different and potentially teaching you in different ways (e.g., a supportive instructor that rewards you when you communicate well versus a punitive instructor that yells whenever you make a safety error).

3.3 Flow

An arrangement of situations – dictated by the events that start them and the order in which they arise – forms a flow of instruction within the game. VITAL allows the user to view this flow directly. Figure 5 illustrates the

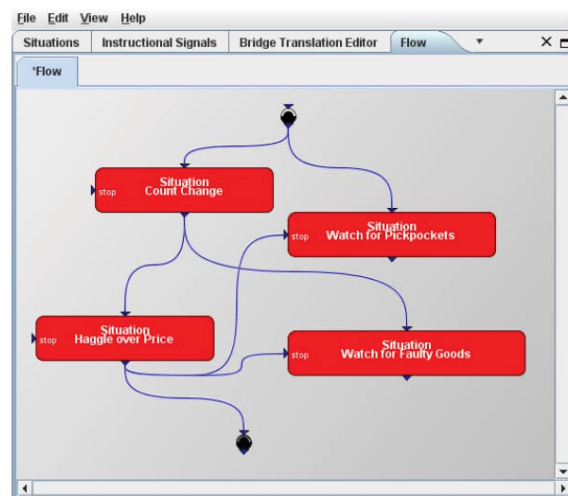


Figure 5. Situation flow within a scenario

situations and the flow among them for a simple game that teaches how to navigate a market in a foreign country. The game is comprised of four situations, one that occurs throughout the game (*Situation Watch For Pickpockets*) and one that starts at the beginning of the game (*Count Change*) and leads to two subsequent parallel situations (*Haggles over Price* and *Watch for Faulty Goods*). This flow captures the instructional sequencing choices. *Count Change* teaches the learning objective *Understand Currency System*, while *Haggles over Price* teaches *Learn Local Bargaining Customs*. The flow we have chosen reflects that we will teach them *Understand Currency System* before *Learn Local Bargaining Customs*.

3.4 Cause-Effect Authoring

To support authoring of instructional logic by practitioners from both instructional and gaming disciplines, VITAL introduces the simplifying concept of ‘cause-effect’ authoring. The flow of instructional actions within VITAL is specified by defining (for each action) what events cause it to happen, and what effects it produces as a result of happening. These events and effects are termed ‘Instructional Signals’, and fall into four categories.

1. Domain signals are those provided directly by the domain ontology. For instance, *Door5.OnOpen* is a signal indicating that a specific door instance (i.e., door number 5) has been opened. *Door5.DoOpen* is a signal representing a command to cause a specific effect in the game (i.e., to open the game door object corresponding to

the instance Door 5).

2. Generated signals are those created automatically by VITAL to capture commonly used events related to core VITAL objects, such as the start and end of situations or the game (e.g., the situation in Figure 2 “Starts When” the generated signal *Game.OnStart* is issued and when it ends it issues the generated signal *Sit.Repair Leak start* as an “End Effect”).
3. User-defined primitive signals are those defined directly as text strings by the author. This allows, for example, an author to indicate that one action will produce an arbitrary signal (e.g., “Entered the shop”) and later define another action that will activate when that signal occurs. This approach was chosen to provide an easy-to-use capability for non-programmers to link up different actions according to their instructional design.
4. Finally, user-defined compound signals are those defined by the author through simple logical combinations of two existing signals. VITAL provides an easy-to-use interface (see Figure 6) for specifying these logical combinations through simple actions such as selecting from a list or choosing a radio button.

Using the Instructional Signals editing interface, a user-defined signal may be

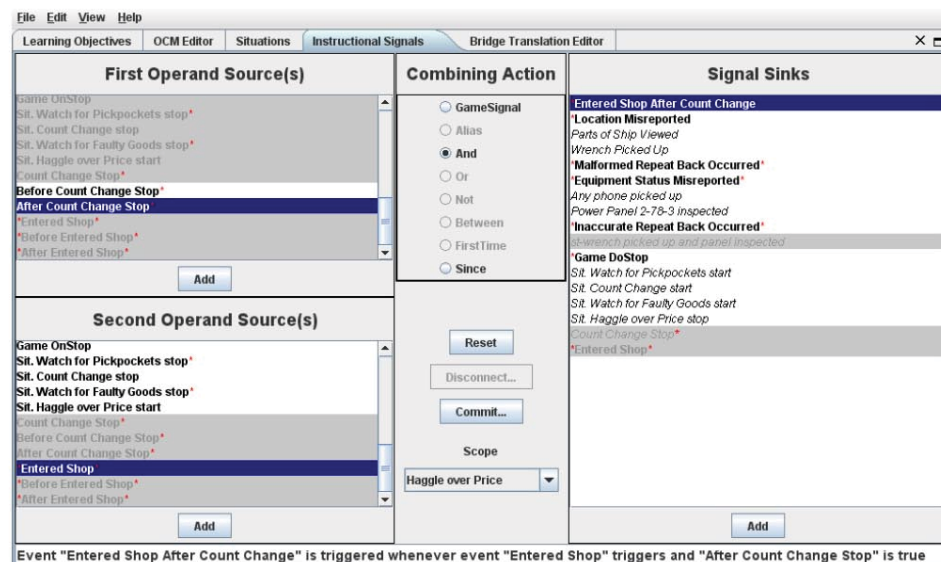


Figure 6. Interface for defining compound instructional signals

directly associated with a domain signal (by selecting the GameSignal radio button). This direct mapping as well as the compound signal capabilities allow high-level authors to create the causes and effects that make sense to them using straightforward wording, and then enable more advanced users to later map those to the domain ontology as appropriate. This flexibility was added to enhance support for non-programmers (e.g., who may have some issues directly using a complex domain ontology).

Cause-effect authoring provides the means to support flexible and complex exploratory learning. In particular, action during the game will typically be driven by the player's choices, and not by a rigid instructional process. For instance, the Situation *Haggle over Price* may be defined to start when the player opens the door to the shop after completing the *Count Change* Situation (see Figure 7). An inspection of the Haggle over Price situation (see Figure 6) shows that it will only start after Count Change ends AND once the player opens the door to a specific room. Thus, it is not just the fact that they have learned how to count change that triggers the subsequent instruction on bargaining (rigid instruction), but the fact that the player has put themselves into a situation where bargaining is now required (flexible instruction supporting exploratory learning). This approach allows the author to define highly complex interactive instruction without requiring a rigid branching structure, and allows the game to provide additional experiences for other (as yet undefined) Situations addressing other (as yet undefined) learning objectives.

Situation Name: Haggle over Price	
Start and End Events	
Starts When	
Entered Shop After Count Change	
Change	

Figure 7. Starting a Situation based on player's action in specific context

3.5 Instructional Mechanics

Situations, cause-effect authoring and domain ontology specification provide a rich array of instructional elements for a game. However, the “meat” is still missing. What is the actual instructional interaction with the student in the game?

To support a rich array of instructional interactions with minimal assumptions regarding instructional methods, VITAL introduces the concept of an *instructional mechanic*. Game designers typically design their games with a focus on the game mechanics that form the game. A game mechanic captures a particular way of interacting with the user – ranging from simple score mechanics or inventory mechanics to complex combat mechanics. VITAL respects this design approach by defining a new class of mechanics. An instructional mechanic (IM) is a game mechanic (or combination of game mechanics) that is used for a particular instructional purpose. The author specifies the instructional logic for a situation by selecting and specifying the details of one or more IMs.

Within VITAL, the author typically selects and specifies an IM using a wizard. The wizard has a number of steps, indicated by the “TO DO” list on the left. For instance, Figure 8 illustrates three steps of a wizard for specifying an instance of a *Time Stoplight Assessor* IM. In the first step of the wizard, the author selects the instructional mechanic type they want to use (e.g., *Time Stoplight Assessor*). The author then gives the IM instance a name and associates it with a specific learning objective. Next, they specify the values for the parameters of the IM. This includes, for example, defining the Instructional Signals that will start and end the IM. In Figure 8, the author has defined this instance of a *Time Stoplight Assessor* to start when the *Report Leak* situation has started (*Sit.Report Leak start*) and to end when the player has successfully completed their report (*Initial Report Completed*). Different IMs have different sets of parameters, and the wizard may include one

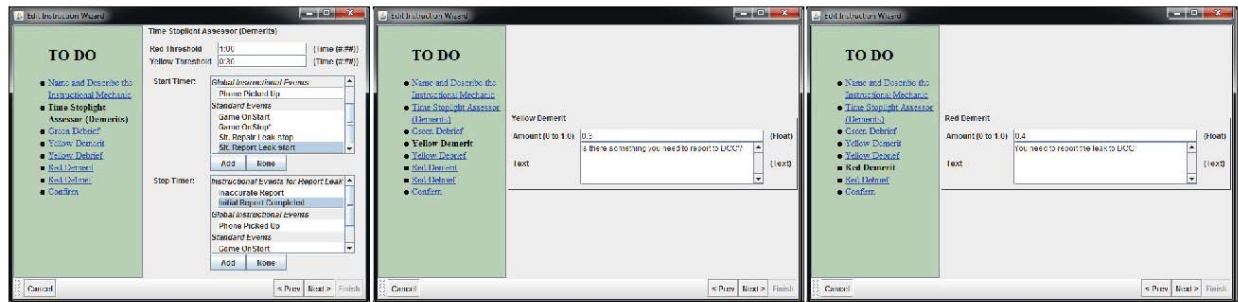


Figure 8. Wizard for specifying an instance of the “Time Spotlight Assessor” Instructional Mechanic

or many steps to specify those parameters. For instance, a *Time Spotlight Assessor* keeps track of how long a player takes to accomplish a particular activity, and moves from green to yellow to red as the player takes too long to perform it. For each level of performance, the author can specify warning feedback and a penalty to give, as well as a message to display during a debrief after the scenario has ended. In Figure 8, the author has specified a penalty of 0.3 (Yellow Demerit Amount) to be given after 30 seconds (Yellow Threshold) and a penalty of 0.4 (Red Demerit Amount) to be given after 1 minute (Red Demerit Threshold) has passed.

VITAL supports a wide variety of IMs, including simple pop-up prompts, complex interactions such as tree-based conversations, and key scenario elements, such as issuing tasks for the player to accomplish. An IM may be used to perform assessment, deliver guidance, provide positive and negative feedback, define the content of an after-action review, and much more. By selecting an appropriate set of IMs and linking them together appropriately using cause-and-effect authoring, a complete set of instructional logic for a given situation can be easily specified. For instance, Figure 2 shows a Situation comprised of 3 IMs. These are listed in the table in the middle of the Situation editing interface. A *Conversation* IM instance named *Make Initial Report by Phone* is initiated when the player picks up the phone (*Phone Picked Up*). Within the conversation, the player may make a variety of choices, some good, others poor. Poor choices result in *Inaccurate Report* signals. These mistakes are in turn counted by a Count Spotlight Assessor IM, which provides

instructional feedback and penalties related to the learning objective *Report Leak Accurately*. Finally, since one of the situation’s objectives is to *Report Promptly Upon Discovery*, a third IM, a *Time Spotlight Assessor*, tracks the amount of time the player takes to make their report and issues feedback and penalties accordingly if the player takes too long.

4.0 EXTENSIBILITY

To support extensibility, most IMs within VITAL are specified using an XML-based data format. The IM specification identifies the underlying game mechanics to use, defines any fixed parameter values of those game mechanics, and identifies the relationships between the game mechanics. To create a new IM within VITAL, a configuration tool is available to define the details of the IM. This configuration is an activity typically performed by a game developer. The beauty of the IM method used by VITAL is that, ultimately, the instructor should not care exactly how the game mechanic is implemented as long as it serves the correct instructional role. Consider a *Guidance Prompt* IM that uses a text-based window game mechanic to display real-time written guidance to the player. In one game, the game mechanic may be a flashy pop-up window in a corner of a screen, while in another it may be an ever-present communication window at the bottom of the screen. This separation of instruction from gaming allows our “instructor” to be specified separately from the “game-master.” It also allows the instructional logic to be transferred from one game engine to another. As long as the new game engine supports the desired

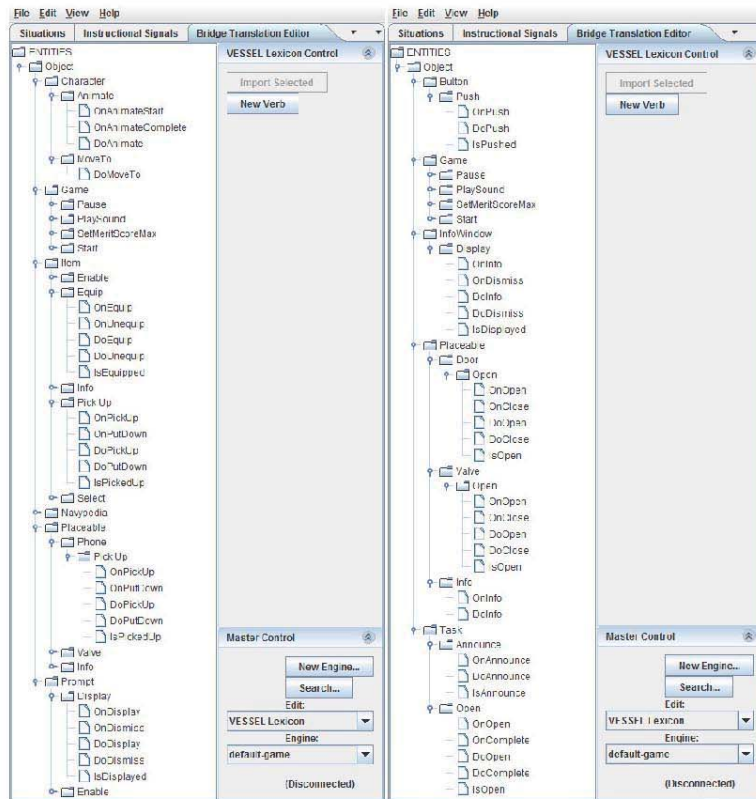


Figure 9. Using VITAL to define different domain ontologies

game mechanics in one form or another, the authored instruction will apply. We have demonstrated the cross-platform transfer of VITAL logic. This demo is available at <http://sourceforge.net/projects/vital>.

As described in section 2.4, a completely new instructional domain is implemented by specifying the ontology of types and verbs instructionally relevant for that domain, as well as a set of instances relevant to the specific material to be conveyed. This provides the capability to apply VITAL to many different domains. Figure 9 illustrates two different domain ontologies that have been defined within the data-driven bridge of VITAL.

To date, VITAL has been integrated with the Delta3D [2], Unity3D [3] and VBS2 [4] game engines. Typically, during development, communication over a standard socket is used (so that the game engine can transact “live” with the VITAL Instruction Editor, for debugging convenience). A separate runtime mode is established for final release (in which

the game interacts with the minimal, VITAL Instruction Engine alone, via XML exported from the Instruction Editor). The VITAL Instruction Engine has been deployed in several different runtime modes, as required to support appropriate integration with different game engines:

- Within Unity3D, the Instruction Engine executes within the main Unity3D process by leveraging a 3rd party tool, IKVM [19] which converts Java to C#.
- Within Delta3D, the Instruction Engine executes in an embedded Java Virtual Machine that is instantiated through C++, in the main game process.
- Within VBS2, the Instruction Engine runs in a standalone process, and a third, specialized process (also standalone) is used to relay messages between VBS2 and the VITAL Instruction Engine.

5.0 CONCLUSIONS

VITAL provides a novel, practical solution to the challenge of authoring of learning games with explicit instructional elements. It provides some initial capabilities for facilitating collaboration between the different disciplines of instructional system design and game development. It also provides powerful middleware for embedding instruction and assessment capabilities in different game engines.

However, there is no free lunch. When starting from scratch developing a new learning game, a great deal of content – both gaming and instructional – needs to be created regardless of the development methods and tools used. VITAL provides a valuable means of creating re-usable components and content that may be used within and across games. However, effort still needs to be made to create the appropriate domain ontology, identify the

required set of instructional mechanics, develop any missing game mechanics in the game, add and configure any missing instructional mechanics within VITAL, and, if using a new game engine not supported by VITAL, build an appropriate bridge to support communication between VITAL and the game engine. In practice, for any given game, there will be VITAL elements that can be re-used and new capabilities that are needed to support the specific needs of the current game. Through an iterative development process, these challenges can be met and overcome. However, close collaboration is needed between the instructional designer and the game developer.

A key issue throughout the development of VITAL was the question “Who’s the user?” Our answer has been “The learning game development team.” This has allowed us to respect the motivations and approaches of both game developers and instructional system designers. However, it made creating a usable tool difficult since the computer-use skill sets of these different professionals are typically very different. One ongoing issue, both with VITAL and future learning game authoring tools, is to identify methods for developing interfaces for a user that, in practice, may not typically exist within a single person.

The VITAL Instruction Engine is available under the GNU Lesser General Public License (LGPL) [20] and the VITAL Instruction Editor is available under the GNU General Public License (GPL) [21].

6.0 ACKNOWLEDGMENTS

VITAL was developed under the Virtual Environments for Ship and Shore Experiential Learning (VESSEL) project, Office of Naval Research contract number N00014-08-C-0030, under the management of Dr. Ray Perez. We would like to thank Ray Tomlinson and Todd Wright for their contributions to the development of VITAL, and the VESSEL project team, including performers from the University of Central Florida, IDEAS, Intelligent Decision Systems, Inc, and Alion Science and Technology, for

their contributions to identifying best practices embodied in VITAL, as well as their efforts to assess and improve the usability of the tool.

7.0 REFERENCES

- [1] Unreal Engine, Epic Games, <http://www.unrealengine.com>
- [2] Delta3D engine, <http://www.delta3d.org>
- [3] Unity3D engine, Unity Technologies, <http://unity3d.com>
- [4] Virtual Battlespace 2 (VBS2), Bohemia Interactive, <http://www.bisimulations.com>
- [5] Simulation, Training and Game Editor (STAGE), <http://www.delta3d.org>
- [6] List of level editors. Wikipedia article. http://en.wikipedia.org/wiki/List_of_level_editors
- [7] Warcraft III World Editor, Blizzard Entertainment, <http://blizzard.com>
- [8] Aurora Toolset for Neverwinter Nights, Bioware, <http://www.bioware.com>
- [9] Maya 3D animation software, Autodesk, <http://usa.autodesk.com/maya>
- [10] 3ds Max 3D modeling software, Autodesk, <http://usa.autodesk.com/3ds-max/>
- [11] Blender 3D content creation suite, <http://www.blender.org>
- [12] Eclipse Integrated Development Environment <http://www.eclipse.org>
- [13] IntelliJ IDEA, JetBrains, <http://www.jetbrains.com>
- [14] Microsoft Visual Studio, <http://microsoft.com>
- [15] Lua scripting language, <http://lua.org>
- [16] Python scripting language, <http://python.org>
- [17] Bugzilla, <http://bugzilla.org>
- [18] GlowCode software profiler, <http://www.glowcode.com>
- [19] IKVM, <http://www.ikvm.net/>
- [20] <http://www.gnu.org/licenses/lgpl.html>
- [21] <http://www.gnu.org/licenses/gpl.html>

5.11 Analog Design for Digital Deployment of a Serious Leadership Game

Analog Design for Digital Deployment of a Serious Leadership Game

Nicholas Maxwell
Booz Allen Hamilton
maxwell_nicholas@bah.com

Tristan Lang
Booz Allen Hamilton
lang_tristan@bah.com

Jeffrey L. Herman
Booz Allen Hamilton
herman_jeffrey@bah.com

Richard Phares
Booz Allen Hamilton
phares_richard@bah.com

Abstract This paper presents the design, development, and user testing of a leadership development simulation. The authors share lessons learned from using a design process for a board game to allow for quick and inexpensive revision cycles during the development of a serious leadership development game. The goal of this leadership simulation is to accelerate the development of leadership capacity in high-potential mid-level managers (GS-15 level) in a federal government agency. Simulation design included a mixed-method needs analysis, using both quantitative and qualitative approaches to determine organizational leadership needs. Eight design iterations were conducted, including three user testing phases. Three re-design iterations followed initial development, enabling game testing as part of comprehensive instructional events. Subsequent design, development and testing processes targeted digital application to a computer- and tablet-based environment. Recommendations include pros and cons of development and learner testing of an initial analog simulation prior to full digital simulation development.

1.0 INTRODUCTION

This paper describes development of a serious leadership game – first as a board game, then as a digital game – and presents lessons learned as to how the “analog first” approach can be an efficient and effective game design methodology.

The Leadership Challenge Simulation (LCS) is a game-based learning experience that helps players to develop their leadership capacity. In the game, players compete within and between teams to acquire effective leadership capabilities and to perform in a series of leadership challenges. Game facilitators guide players through the experience, leading discussions and debriefs focused on understanding the nature of leadership, the difficulty of developing leadership, and the application of lessons learned in the game to real-world situations.

The LCS was first developed as a facilitated board game, which was used as a test environment for later development of a digital game. The board game was piloted and delivered in the initial organization and was then adapted for delivery in different organizations. Evaluation of the analog game experiences informed game revision, which had the benefits of being relatively inexpensive and requiring no hardware, software, or network access. Once game mechanics and facilitation protocol were established, game designers, software developers, and instructional designers worked together to understand the steps to translate the game experience to a digital format, retain the most valuable elements of the board game, and make changes to adapt to a digital game experience.

This paper is presented in three parts. First, we describe the design and development of the initial board game, including how organizational data were analyzed to create a game experience that met precise learning objectives. Second, we describe

how the board game was evaluated and the lessons learned that are relevant to analog as well as digital environments. Third, we describe the process for turning the board game into a digital game, primarily by presenting and explaining the steps taken that can be replicated by others.

The intended benefit of this paper is to provide lessons learned regarding the use of an “analog first” method to designing digital games. These lessons include how to design an analog leadership development game, how to evaluate its design and effectiveness, and how to use an analog design and its evaluation in preparation to create an effective digital game for leadership development.

2.0 CLIENT NEEDS AND OBJECTIVES: OVERALL PURPOSE

2.1 Organizational needs analysis

The LCS was first conceived as a proactive way to address identified future challenges in organizational leadership in one specific organization. The organization, referred to in this paper as “FerroTech,” faced age-related demographic risks. Specifically, about two-thirds of their workforce was eligible for retirement within the coming five years. Many of these retirements would come from the leadership ranks of the organization. The organization had identified thousands of high-potential future leaders to fill their ranks, through performance appraisal data and manager ratings on readiness for promotion. However, many of these candidates with high potential were rated as lacking the leadership skills that were essential at higher levels of leadership, suggesting a future cadre of leaders with potential, but not enough leadership skill to perform as the organization would need.



Figure 2.1. Analog Game Testing

3.0 GAME

3.1 The Case for a Game for Leadership Development

We incorporated the application of a “serious game” to the problem of rapidly training candidates with high potential in the various aspects of leadership that they will be expected to display. By engaging candidates in interesting and active media, FerroTech has its best opportunity to increase the awareness, understanding, and retention of these various leadership qualities.

Games used to simulate real-world dilemmas fall into a class of gaming known as “Serious Games.” These dilemmas may range from governance and stability issues to terrorism to developing leadership capabilities. Serious games are an interactive, highly customized, and attractive means of transmitting content to a broad audience. In particular, they are a powerful medium for simulating situations and posing questions that a person would not normally encounter. They can also foster competition, which can also act as a spark for social interaction. In this manner students become fully immersed in the experience and engage as co-creators of the learning experience, in contrast to

traditional, presentation-heavy learning techniques that lose the attention of students and fail to deliver lasting impact [1].

Many of the reasons to use an experiential-based game as the primary training medium lie in adult learning theory. The research into adult learning is quite clear that adults learn differently than what is supported by a presentation/lecture-only university model [2]. Adults need to understand why they should learn. They are motivated to learn as they experience needs and interests that learning will satisfy. Adults need to be able to understand the value in doing what they are being asked to do. They also have a strong need to be self-directed. They know what they want to learn and want to do with newfound knowledge.

The orientation of adult learning is life-centered, such that they learn best when they understand how the content of the learning event will fit with their life experiences. Immersive experiences are thus the richest resource for adult learning; they provide experiences in which adults create their own interpretations and attach their own meaning to events tied directly to the learning objectives.

Feedback is a second critical element of adult learning that is central to serious games. Immersion in an experience alone does not guarantee learning. Adults learn by sharing their experiences and engaging with various perspectives to produce richer interpretations of experience. By understanding the experience of fellow students on a particular issue, adults can place themselves into situations they have not yet experienced. Effective serious games drive precisely this combination of new first-hand experiences with rich feedback to draw out valuable lessons and connect them to their own life experience.

3.2 How to Design a Board Game for Leadership Development

At the base level, the design of a board game is arguably the most critical aspect in the development of a program to cultivate leadership qualities. A properly designed game will identify proper issues, analyze the best way to present them to the user, and provide quantifiable results so that the objectives of the program may be met.

The first step in quality game design identifies the objective. In the case of FerroTech, we understand that the purpose of this game must be to rapidly bring high-potential candidates to a level where they can acquire leadership skills necessary to succeed at the next level within their organization. The second step is to identify the issues that need to be addressed. This includes understanding the competencies and qualities that successful leaders have at FerroTech, as well as how they developed them. Based on a solid understanding of these leadership qualities, game designers can then work to postulate the desired output, which is the specific way to demonstrate successful results to the client.

Once a good foundation has been established in understanding the purpose of the game and the issues to be addressed, game designers can focus more on the game itself. This will include developing the game scenario and the game environment. Next, developing the game structure will allow designers to identify the equities to be addressed, while also incorporating creative tension in the game. This is essentially the quality of game design that keeps participants interested in the game. After this has been developed, the designer may identify the timing of the game and the moves to best address the objectives, after which he may actually prepare game materials and models.

The entire game design process comes down to developing something that correctly addresses the topic of interest, while also

Table 3.2. “Analog First” Process Phases

	Phase	Lessons Learned
1.	Identify the Objective	It is important that the analog game being developed will fulfill the needs of the client. The objective of the LCS was to rapidly bring high-potential candidates to a level where they can acquire leadership skills necessary to succeed at the next level within their organization.
2.	Identify Issues to Address	With the objective in mind, this phase requires analog game designers to figure out the key elements involved in accomplishing the client's mission. The LCS sought to understand the competencies and qualities that successful leaders at FerroTech have and how these were developed.
3.	Develop Analog Game Scenario and Environment	Once key elements and the objective are known, it is the job of the analog game designers to come up with interesting and enjoyable ways for these elements to interact in an analog game.
4.	Develop Analog Game Structure	Deciding what and where to place elements on the analog game board as well as creating intuitive interactions with game pieces will require careful thought during this phase.
5.	Design Iterations with “After-Action Reports”	This phase involves working with the client to make sure that the analog game is on par with the client's expectations. Collecting valuable feedback will allow game designers to return to the phases above to revise design decisions.

keeping players interested. In order to make players aware of the leadership qualities that they should be developing, it is essential for game designers to prepare an “after-action report” to identify the results of the game and next steps for game facilitators and FerroTech training administrators.

3.3 Facilitation Design to Maximize Learning

Just as the research shows that sharing experiences is not enough, the research points to the role of the “faculty” in an adult learning environment. The evolving roles are modeling (in which the master demonstrates behavior), coaching (in which the learner does the complex task with the master observing), and fading (where the master only gives help when needed) [3].

The LCS makes use of a facilitator in each of these three capacities to maximize the benefit to the game participants.

Design and facilitation of the LCS includes an additional focus on experiential variety as well as feedforward (not only feedback) in order to develop adaptive leadership capacity. Facilitators guide learners in adaptive leadership behaviors, specifically metacognition and self-regulation, to help learners practice and refine skills necessary to overcome the variety of challenges in the game. This approach is grounded in recent research in leader adaptability and responds to a call for greater attention to adaptive versus routine expertise in training and development [4].

4.0 GAME EVALUATION

Before moving the LCS into the digital space, a few design concerns were raised involving the abilities and disabilities that digital games have compared to analog games. This section describes these concerns.

4.1 What a Digital Game Can Do

Bringing an analog board game to the digital stage opens up design opportunities and game enhancements that are not easily feasible for analog games. Consider the fact that analog board game designers are limited by the amount of physical board space that is available to them. For example, an analog designer may be restricted to keeping a board to the size of a standard dinner table. This is not the case for a digital version of the same game. With a well-planned graphical user interface (GUI), digital designers are able to present much more information on a single computer screen than a dinner table-sized physical board [5]. On the LCS board, a chart that describes pre-defined development behaviors for a number of required tasks takes up much of the physical board space. During the analog design of the game, the placement of this chart was under scrutiny and the number of possible development behaviors was limited. For a digital design of the LCS, there is no longer a board space dilemma. With the ability to page and scroll data on a computer screen, the digital re-design of the LCS for organizations other than FerroTech that may identify many more development behaviors, would not run into a game-changing decision when it comes to physical board space.

As another consideration, the ability to have a board and pieces on a board appear differently to each player is a considerable challenge for a physical, analog board game. This effect can be achieved in a variety of manners on a physical board although the methods seem a bit haphazard

when compared to the style in which it would be implemented in a digital version of the same game. In a digital version, each player can be given an individual device in which a centralized server would manage the visualizations that each player is able to see on their device's screen. This implementation allows digital game designers to easily comply with fog of war rules, such as allowing players to see only up to three spaces beyond their current position on the board, or not allowing players to see opposing player's pawns or resources. During the analog design of the LCS, there was a question of how many tasks would be visible to users on the development behaviors to tasks chart mentioned above. Analog designers considered covering up the tasks that were not available to users in the current stage of the game as a fog of war technique. This feature would have been easily implementable in the analog version by covering up columns on the development behavior to tasks chart, although this would have led to further intervention by game mediators, and another dynamic component to an analog game that many users had already perceived as "complicated." In a digital version of the game, this feature can be implemented in a similar manner, by covering up columns on the development behaviors to task chart, although this would be mediated automatically by a central server. A digital design also offers distinct opportunities to reward players for good performance. For instance, players that are first to circle the board can be given the ability to view additional columns on the development behaviors to task chart, or the ability to see further than the fog of war limit.

Further enhancements that a digital game allows that are not easily implemented in analog games include the ability to display real-time probabilities and statistics as well as post-game analysis results of a game. In a digital version of the LCS, before players roll dice, they may be presented with the probabilities of landing on a tasking board space as well as the probabilities that they

will land on a tasking space if they choose to trade in for more dice as a decision aid. Furthermore, a digital version of the game can easily allow the events of a game to be tracked and stored as a linear series. This allows the ability to display real-time and post-game statistics and also allows game moves to be undone easily during a game or replayed during a post-game analysis.

4.2 What a Digital Game Cannot Do

Although a digital version of a game may provide various enhancements to an analog game, there are a number of characteristics that a digital game cannot simulate as well as an analog version of a game such as the LCS. For one, players may wish to communicate and interact with one another to discuss game strategy or to ask for advice, and doing so in a digital environment may not seem as natural to them. Much study has gone into assessing the manners and abilities in which individuals communicate in virtual environments as opposed to physical environments [6]. These studies have shown that certain virtual environments are able to reproduce many communication aspects that are present in physical environments although there are characteristics of physical environments that are difficult to reproduce in a digital environment. For example, individuals rely heavily on body language and gestures in physical communication. Digital environments that attempt to display or recreate these gestures often require plenty of costly network bandwidth in order to display a live video feed or a physically mapped avatar.

Another advantage that an analog version of a game such as the LCS has over a digital version of the game is the fact that these analog games are often times designed to be facilitated by mediators. These mediators are typically experts in the simulation field and are trained to mentor and coach players throughout the course of a game. A digital version of the LCS would

likely mitigate much of the interaction that players would have with mediators, which may reduce the ability for players to absorb many of the messages that the game is designed to engrain in its players.

The vigorous involvement that players experience as part of an analog board game plays a key role in a player's understanding of a game, its rules, and any learning points that a game is designed to teach players. For instance, in the LCS, if a player wishes to develop a leadership behavior, they must pay one resource into a resource well. The act of physically taking a resource chip and placing it in the resource well better reinforces the concept that it costs to develop a leadership behavior, as opposed to a computer application taking care of this step automatically.

5.0 DIGITIZATION PROCESS AND NEXT STEPS

After assessing the possible enhancements as well as the limitations of a digital version of the LCS described above, we started developing a digitization process. To start the process, we needed to describe the game as a state machine in which players can perform actions to modify the state of a game. To do this, we divided the LCS into its basic components and described each in a generic sense. We were able to describe the analog board as a set of nodes and edges in a directed graph. The characteristics of game pieces were then described, including all pawns, chips, and dice. Finally, we decided on the set of actions that players make during the course of a game. Describing the game, in detail, in this generic way allows a software development team to begin implementing the game logic.

The next step to full digitization would be to take these components and create a digital model of the initial state of components that can be stored and read from a data store. After accomplishing this step, a digital

development team could work in parallel to describe the rules as results of the identified actions that players may take, as well as take the digital information and graphically present this to a user.

6.0 CONCLUSION

After the digitization process, we came to realize that the “analog first” methodology was a very effective way of developing a well-planned leadership development game that players will eventually be able to play in a digital environment. Table 6.1 below describes the “analog first” process phases and the key lessons learned for each phase.

After a playable analog game has been developed and incrementally tested for consistency with the client's objectives, addresses key issues important to the client, and is enjoyable to play, the game is ready for the digitization process. The digitization process includes describing the analog game in its most basic components, followed by identifying the set of actions that players make during the course of the game. After an analog game has been fully digitized, further enhancements to the game may be incrementally built in to a digital version.

7.0 REFERENCES

1. Salas, E., Wildman, J. L., & Piccolo, R. F. (2009). Using Simulation-Based Training to Enhance Management Education. *Academy of Management Learning & Education*, 8(4), 559-573.
2. Knowles, Malcolm S. *The Adult Learner: a Neglected Species*. Houston: Gulf Pub., Book Division, 1984. Print.
3. Collins, A., Brown, J.S., & Newman, S.E. “Cognitive Apprenticeship” In L. B. Resnick (Ed.) *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, pp. 453-494, Hillsdale, NJ: Lawrence Erlbaum Associates, 1989.
4. Nelson, J. K., Zaccaro, S. J., & Herman, J. L. (2010). Strategic information provision and experiential variety as tools for developing adaptive leadership skills. *Consulting Psychology Journal: Practice and Research*, 62(2), 131-142.
5. Ware, Colin. *Information Visualization Perception for Design*. Amsterdam: Elsevier, 2004. Print.
6. Rekimoto, J., Ayatsuka, Y., Hayashi, K. (Oct 19-20, 1998). "Augment-able reality: Situated communication through physical and digital spaces," *Digest of Papers: Second International Symposium on Wearable Computers*, 68-75.

5.12 Zero-Fidelity Simulation: Engaging Team Coordination without Physical, Functional, or Psychological Re-Creation

Zero-Fidelity Simulation: Engaging Team Coordination without Physical, Functional, or Psychological Re-Creation

Zachary O. Touns

Crisis Response Innovative Technologies Lab
Disaster Preparedness & Response, Texas Engineering Extension Service
+ Texas Center for Applied Technology
+ Interface Ecology Lab | Texas A&M University
zachary.touns@teexmail.tamu.edu

William A. Hamilton
Interface Ecology Lab | Texas A&M University
bill@ecologylab.net

Andruid Kerne
Interface Ecology Lab | Texas A&M University
andruid@ecologylab.net

Abstract. Team coordination is essential across domains, enabling efficiency and safety. As technology improves, our temptation is to simulate with ever-higher fidelity, by making simulators re-create reality through their physical interfaces, functionality, and by making participants believe they are undertaking the simulated task. However, high-fidelity simulations often miss salient human-human work practices. We introduce the concept of zero-fidelity simulation (ZFS), a move away from literal high-fidelity mimesis of the concrete environment. ZFS alternatively models cooperation and communication as the basis of simulation. The ZFS Team Coordination Game (TeC) is developed from observation of fire emergency response work practice. We identify ways in which team members are mutually dependent on one another for information, and use these as the basis for the ZFS game design. The design creates a need for cooperation by restricting individual activity and requiring communication. The present research analyzes the design of interdependence in the validated ZFS TeC game. We successfully simulate interdependence between roles in emergency response without simulating the concrete environment.

NOMENCLATURE

FER – Fire Emergency Responder; any trained response professional dispatched for firefighting, rescue, command, etc.

IC – Incident Commander; the person responsible for coordinating response to an emergency incident, as defined by the Incident Command System.

TeC – The Team Coordination [Game]; a zero-fidelity simulation of the information- and human-centered aspects of teams in fire emergency response work.

1.0 INTRODUCTION

Team coordination is an essential aspect of work contexts in which tasks are too large or complex for one person. In the emergency response domain, such work must be performed quickly in a dangerous, high-stress environment. Coordination consists of cooperation and communication within a team to accomplish a set of shared objectives.

The present research begins with an investigation of team coordination in fire emergency response work practice [21]. To address a discovered need for team coordination education, we developed a game in which players exercise cooperation and communication, rather than addressing the whole firefighting experience. In this process, we abstracted out firefighting's varied perspectives, sources of information, and communication. We situated them within a stressful game environment. We hypothesize that the result, while barely recognizable as a simulation, constitutes a *zero-fidelity* simulation (ZFS) of team coordination. The ZFS Team Coordination Game (TeC) offers a focused educational experience that our research has shown to be effective for engaging emergency responders in team coordination [19][20]. By design, it eschews many of the typical notions of simulator fidelity: physical, functional, and psychological.

Game mechanics are the activities that players undertake in a game [14]. They are experiential building blocks in which a player

makes a decision and the game responds with a change in state. A *core game mechanic* is a mechanic that is repeated during play.

In this paper, we present TeC core mechanics and connect them to practice. We show how the interdependency between players engages them in processes of communication like that of emergency responders. Evidence comes from a series of user studies of TeC using fire emergency response students. Details of the studies can be found in [19][20].

In the following sections, we address background in team coordination, including details of fire emergency response work practice. We define zero-fidelity simulation in the context of simulation in general, then the Team Coordination Game in specific. We cover, in-depth, the interdependency of action and information in the game, beginning with a demonstrative scenario taken from user study data. We conclude by discussing the implications of ZFSes and our design considerations in successfully building a ZFS.

2.0 TEAM COORDINATION

Team coordination is characterized by the synchronization of action by multiple individuals to accomplish a task. Cooperative action and communication are essential. Team effectiveness and efficiency is improved if team members are skilled at coordinating.

Individuals build and maintain *mental models*, cognitive structures that support problem solving by predicting future outcomes [6]. When *shared* [10], mental models enable team members to predict each other's needs, informationally and practically. This leads to *implicit coordination*, collaborating with reduced communication [5]. Implicit coordination theory posits that communication is costly, in terms of time, cognitive, and technological bandwidth. Conserving bandwidth, for example by providing needed information without being asked, is thus beneficial [11].

Distributed cognition theory accounts for the ways that information is split out among individuals, artifacts, and time [8]. It addresses

the ways in which individuals transform information across media to make use of it. Distributed cognition forms the basis for our understanding of information-centric team tasks and informs the design of TeC.

2.1 Fire Emergency Response

The present research is based on fire emergency response work practice. We began with an ethnographic investigation at one of the world's largest firefighter training schools, Brayton Field in College Station, Texas, USA [21]. We interviewed expert responders and observed high-fidelity simulation burn training exercises, in which students fight actual fires.

Fire emergency response is accomplished by distributing a number of small teams in and around the fireground. An incident commander (IC) directs the response from a distance, possibly consulting information artifacts. The IC relies on reports from deployed responders, as well as observations of the incident as a whole. In this context, information is distributed among team members, and each team must develop situation awareness by integrating first-hand and communicated data. The IC constructs plans for the deployed teams, while the teams gather and share information and use it to take situated action [17] in the context of the plan. *Information is distributed* throughout the team, and resolving and sharing information is central to response. While responders prefer to communicate face-to-face, radio is frequently necessary to connect teams that are distributed. Radio is problematic, because it is half-duplex and difficult to understand.

Our results indicated a successful team coordination simulation should capture the ways that information is distributed and force participants to mix communication modalities (face-to-face, radio) [21]. Based on our observations, it was clear that there was a need for educating communication under stress, but no need to directly mimic the context of firefighting. Students needed to learn to implicitly coordinate in a distributed cognition environment.

3.0 ZERO-FIDELITY SIMULATION

Simulations are operational environments that facilitate learning and practicing; they are typically defined in terms of *fidelity*, the degree to which a simulation recreates some aspect of a real environment and activity. Fidelity is multi-dimensional; a broad classification identifies how a simulation re-creates a target environment physically, functionally, or psychologically [1][2][13]. Physical fidelity addresses the way in which the simulator resembles the target environment, while functional fidelity is the degree to which the environment reacts appropriately. In a flight simulator, physical fidelity addresses the controls and gauge configuration, while functional fidelity addresses the way those gauges and the view from the cockpit change in response to the controls. Psychological fidelity identifies the degree to which a participant believes they are undertaking a task in the target environment.

While the temptation is to build systems with ever-higher fidelity, assuming automatic skill transfer, we take a different tack: *we do not re-create the target environment at all*. A number of approaches address high-fidelity

simulation, employing complex algorithms and graphics processing power to construct a realistic virtual environment [16][18]. These achieve a high level of physical and functional fidelity with the assumption that psychological fidelity follows [7]. Even in low-fidelity simulations, high psychological fidelity is essential; participants should *believe* that they are working on a particular task, despite the simulator not completely recreating the task environment [3][9]. *Zero-fidelity simulation* is different in that the suspension of disbelief is not required to effectively participate in the simulation, yet educational value is retained.

Zero-fidelity simulation engages the participant in the information- and human-centered aspects of the simulated activity. Tasks are then re-situated in an alternative environment. The ZFS Team Coordination Game that we present does not encourage players to believe that they are fighting fire, nor does it address the physical activity of firefighting, yet we have shown that FER students engage in and improve in the same communication and team coordination tasks required in fire emergency response and carry their skills forward into

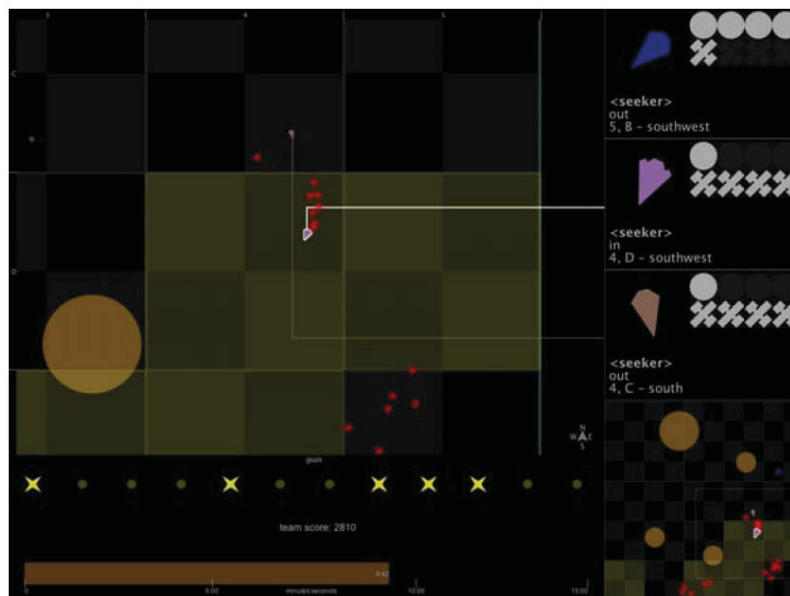


Figure 1. Coordinator game interface. The main view shows terrain with all entities in it; yellow regions contain goals. Walls (visible to seekers) cannot be seen. The right column shows the status of each seeker; a mini-map is at the bottom. Below the main view is the list of goals (here, five remain un-collected), team score, and the time remaining.

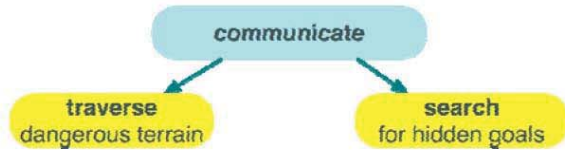


Figure 2. Core mechanics of TeC. Traversing and searching and dependent on communication.

high-fidelity burn training exercises [19][20].

4.0 ZFS TEAM COORDINATION GAME

The Team Coordination Game is engineered to simulate the distributed cognition environment of fire emergency response. Participants are mutually dependent on one another for information. Here, we present TeC's information distribution across roles, terrain and game entities, core mechanics, and motivation through score.

4.1 Roles & Information Distribution

Players take on the role of the *coordinator* or a *seeker*. These roles are differentiated by the player's agency in the game world and are the axis along which information is distributed.

Players in the seeker role control a seeker *avatar* in the game environment. They have a high-detail view of the terrain around them (Figure 3). Some information is specifically hidden and is only available to the coordinator.

The coordinator cannot directly act in the game world, rather s/he must communicate with the seekers to manage information and develop strategy. In this way, the coordinator role is like that of an IC. The coordinator has access to low-detail information about the entire game world (Figure 1).

To engage players in mixing communication modalities [21], games are run in one of two conditions: co-located or distributed. In the co-located condition, all seekers play around a table, while the coordinator is isolated and may only be reached using a half-duplex radio. In the distributed condition, all players must use the radio to communicate. In some cases, players can use fast, face-to-face communication, while in others they must use the slow,

lossy radio. This leads to careful decision making about which modality to use.

4.2 Terrain & Game Entities

TeC game play is made up of interactions between game entities. Seekers drive *avatars*. *Threats* chase avatars and take them out of play, preventing them from collecting *goals*, the main objective. Game *terrain* limits how seekers' avatars move in the simulation.

Game terrain consists of a number of elements that influence play. Walls impede avatars in the virtual environment; walls are visible to seekers, but not the coordinator. Offline regions of the terrain make a seeker invisible to the coordinator as well as threats, but prevent the seeker from collecting goals. Bases are safe for seekers, but are invisible to them; such information is available only to the coordinator.

Seeker players act in the simulation by controlling avatars. Seekers may move in terrain, observe their local surroundings in high detail, and search out and collect goals. Seeker avatars have *hit points* (HP). When a seeker avatar has HP the seeker is *in*, if the seeker avatar has 0 HP then it is *out* and unable to succeed at collecting goals. To regain HP, seekers must wait in a base.

Threats protect *goals* and pursue seekers in the game world; they represent danger (Figure 3, a., c., d., e.). When in contact with a threat, seekers lose HP. Threats are controlled by a continuous flocking simulation [12]. Within this simulation, threats patrol and guard goals and also pursue nearby seekers. Seekers that are collecting goals attract more threats. Threats do not pursue offline or out seekers.

Seekers collect goals in the environment. To collect a goal, a seeker must either point his/her avatar at the goal or move on top of it for a brief period. Seekers must be online and in to collect. Cooperative goals require multiple seekers to collect the goal simultaneously. A goal indicates how many seekers are required by the number of rings around it

(Figure 3, a., b., d., e.). These rings fill with the collecting seekers' colors to indicate progression (Figure 3, b., d., e.).

4.3 Core Game Mechanics

TeC is built up from three core mechanics: traversing dangerous terrain, searching, and communicating (Figure 2). In practice, each of these is a suite of activities executed in coordination by players (Figure 4). To succeed, seekers must traverse dangerous terrain to search for goals, while the whole team engages in communication.

4.4 Motivation Through Score

Players are motivated to engage in TeC's core mechanics and increase performance through score. Score in TeC is shared by the entire team and is computed in such a way as to motivate the core mechanics and team coordination. The team receives points for being in dangerous terrain, for collecting goals, and for completing a game faster than the allotted time. Collecting cooperative goals nets significantly more points than individual goals.

Points are deducted when a seeker is taken out. Also, no points are awarded if the seeker is not in danger, discouraging players from hiding in bases or offline areas.

5.0 INTERDEPENDENCE BY DESIGN

In this section, we look at the three core mechanics of TeC. Seekers engage in traversal and searching. Each of these core mechanics is a suite of other game mechanics, and each of those mechanics is dependent on information that may or may not be available to any one seeker.

When the necessary information is not available to a particular player, that player must engage in the third core mechanic: communication. Communication is one of the most critical components of team coordination and the most important core mechanic. The designed information distribution engages players in communication in service to the other two core mechanics. We connect the core mechanics to observed activities in fire emergency response work practice. We begin with a



Figure 3. Simulation scenario from user study data demonstrating core mechanics. Seekers traverse terrain and search for goals while communicating.

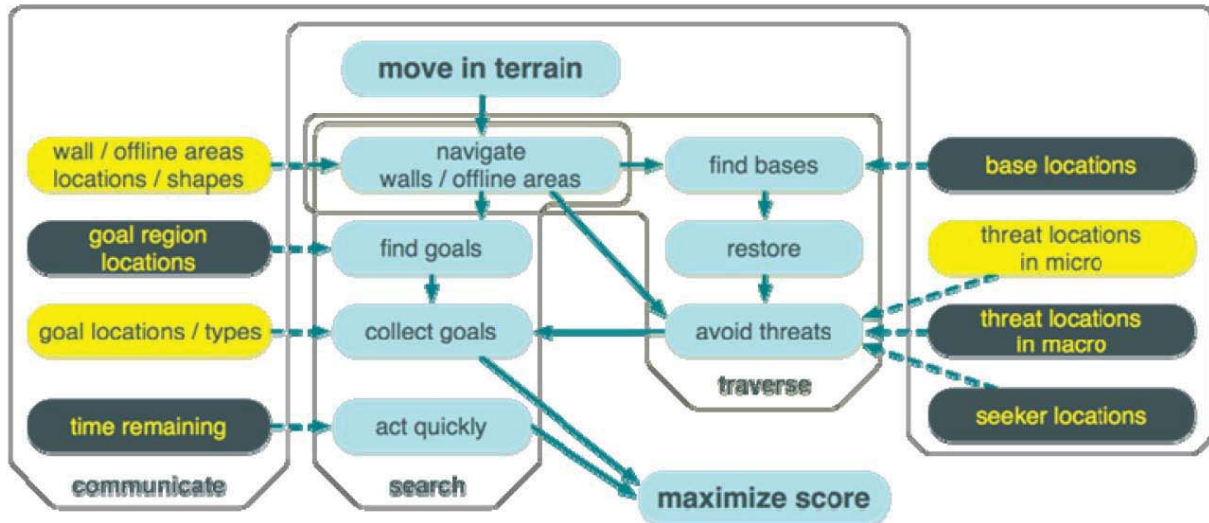


Figure 4. Interdependency of game mechanics and information sources in TeC. Seekers engage with game mechanics (light blue); game mechanics that are pre-requisite to others are shown with an arrow. Information sources come from a seeker's own interface, the interfaces of other seekers, and the coordinator. Communication is essential to get timely information and maximize score. Seeker information sources are in yellow; coordinator sources are in dark gray. Core mechanics that encompass game mechanics and information sources are outlined.

scenario taken from user study data with fire emergency response students that illustrates the mechanics in action. Figure 4 diagrams game mechanics and their interdependencies with information sources, situating these with the core mechanics of traverse, search, and communicate.

5.1 Scenario

The present scenario is an account from actual game play data from a user study of FER students enrolled at Brayton Field [19]. Team calTEXANADA is nearing ten minutes left in their fifth TeC game. As it will turn out, three of the team members will be in the top 10 of their class at the Firefighter Training Academy (2nd, 4th, and 6th). By now, they are starting to play effectively as a team.

In this game, the three seekers are seated around a table at notebook computers, facing one another (co-located condition). The fourth team member, the coordinator, observes the simulation world from another room and communicates with the seekers using a half-duplex radio.

At the beginning of this scenario, the coordinator calls to the seekers and informs them that there is a goal nearby; one seeker acknowledges for the team. With some searching and traversal around the terrain, [Seeker 1] spots a goal that will require cooperation by all three seekers to collect (Figure 3, a.). The move is risky: a threat is already closing on [Seeker 1] and collecting the goal will subject all three seekers to attack by that threat.

The seekers rush the goal as the threat comes closer. Two successfully start to collect it, but [Seeker 2] does not line up correctly (Figure 3, b.). A brief conversation resolves the problem. [Seeker 2] adjusts his position, and the three run away from the threat before it can attack.

[Seeker 2], now pursued, quickly tells his co-located teammates he is "...headed south! I'm headed south! I'm headed south!" (Figure 3, c.), urgently keeping them apprised of his escape plan. [Seeker 1] responds to tell him they are following; the team (and threat) follow [Seeker 2].

The coordinator implicitly coordinates, sharing information that only he has: there is a base

nearby. [Seeker 2], however, has found another cooperative goal during his flight (Figure 3, d.). Having sustained only a little HP damage from the threat, he is still in and rushes the two-seeker goal while notifying his teammates of the information in his local scope and outside of theirs. [Seeker 1] and [Seeker 3] are far behind as another threat joins in the pursuit.

[Seeker 2] stays near the goal while being attacked by two threats. [Seeker 3] rushes through quickly and assists in collecting the goal (Figure 3, e.). All the seekers run away to safety before they are taken out completely by the threats (Figure 3, f.).

Throughout the scenario, the seekers evoke the core game mechanics while under stress: communicating, traversing dangerous terrain, and searching for goals. The seekers make use of rapid changes in information in their interfaces to respond to a failed goal-collection attempt. They opt to speak to each other face-to-face, instead of contacting their coordinator over radio. We also note that the coordinator performs implicit coordination when he notifies the seekers of a nearby base without them asking for it.

5.2 Traversing Dangerous Terrain

Like firefighters, seekers traverse dangerous terrain, *about which they have only partial information*. At a structure fire, firefighters and the IC must cooperate and communicate to construct a satisficing [15] image of the incident status. Firefighters move room-to-room, with limited visibility, to search for victims and develop a sense of the tactical details of the incident [4][21]. Important discoveries are relayed to other team members and acted upon. The environment consists of a number of health hazards, the most obvious of which are fire and structural collapse. The IC, safely located away, combines the perspectives from the various workers and uses contextualized external observations to make decisions and strategically direct the firefighting teams. Each part must combine perspectives from the others to succeed.

In TeC, seekers traverse a virtual environment that contains walls and threats while under a time limit. The threats and time limit create a stressful environment, while the walls impede progress. Unlike firefighting, threats do not behave like fire or other hazards and they cannot be defeated. As in firefighting, each part of the team has access to different pieces of the information picture. Seekers are aware of local walls and threats, but the coordinator is aware of time remaining, safe bases, and threats in macro. While under stress, seekers must communicate about navigation and rely on the coordinator to provide an overview and direct them strategically.

TeC simulates the information interdependence of the task of traversing dangerous terrain from firefighting. Firefighters and seekers traverse terrain, avoiding hazards and circumnavigating obstacles to search the environment. The IC / coordinator, isolated from those on the scene, develops strategy and communicates macro information about the situation, building on information provided by those in the field.

5.3 Searching for Hidden Goals

Searching physical environments involves plotting a careful and situationally informed route through a structure. Searches may be coordinated with help from the IC, who may have access to plans and who will provide strategic advice on where to search. Searches are also coordinated within the deployed teams, as the various perspectives and locations of firefighters are important. It is a waste of time, energy, and possibly lives to search an area multiple times, and so status reports are necessary. It is essential that the search be complete, that all victims be found, and that all fires are either suppressed or discovered and left to burn.

In TeC, seekers move through the game terrain to locate goals. To optimize score, seekers must find all goals in the minimum amount of time possible. The coordinator has an overview that includes the regions that contain goals, but no specific details. This enables the coordinator to direct seekers in macro, but not

in micro. The micro situated actions of seekers are based on information that they, the seekers, have. This results in seeker-seeker information distribution.

TeC simulates the same individual-individual information distribution found in the structural search task. The alternate perspectives of seekers enable them to build up collective situation awareness and use distributed cognition to their advantage. This process is a simulation of the combining of individual perspectives required in structural search.

5.4 Communication

Communication is the most important core mechanic of TeC. Both traversal and searching rely on distributed information, thus requiring the use of communication. Similarly, in firefighting, team members must share information constantly to build situation awareness.

In our zero-fidelity simulation, we make use of half-duplex radios, like those FERs use. Such devices allow only one participant to broadcast at a time, meaning that bandwidth needs to be used carefully and thoughtfully. These considerations are thus brought into play by retaining the radio technology for communication.

6.0 DISCUSSION & CONCLUSION

Many of the interdependencies in TeC are analogs of the work practice of firefighting, but abstracted. TeC is specifically not themed, it is generic by design. This enables players to build their own interpretations of the game environment, some players, for example, called threats "bad guys" and others, despite the design, called them "fire" [20]. One team, Team Firestorm, noted that the threats were not at all like fire, but the stress they created encouraged them to communicate quickly [20]. The lack of theme opens the potential to use TeC in other domains.

All TeC core mechanics are information-centric. Players need information about the game environment and one another to make decisions about which game mechanic to engage when and how to parameterize such ac-

tions. Interdependency derives from the reliance on different team members to provide data about their own perspectives. The result is a need to engage in the core mechanic of communication.

TeC players learn team coordination skills [19][21]. Despite deriving from fire emergency response and building on the roles of firefighter and IC, players do *not* directly learn about firefighting, fireground tactics, or incident command. Rather, they learn how to make quick decisions about what information is needed by teammates, how best to communicate that information, and when it is essential. They learn to cooperate and communicate while under stress.

As a zero-fidelity simulation, the Team Coordination Game does *not* directly connect to the concrete environment. The tasks performed are *not* directly like those of fire emergency response, but the human-human interaction requirements are simulated. The interdependency between team members requires that they communicate efficiently to succeed. Score serves to motivate interaction with the core mechanics. The implication is that psychological re-creation, like physical and functional, is neither *retained* nor *needed to teach team coordination*. Other domains and tasks may benefit from this discovery, which indicates that expensive high-fidelity is not always the correct educational tool. We hypothesize that high-fidelity may even be distracting to learners, who, in expecting a replication of reality, may fixate on flaws in the simulation.

ACKNOWLEDGEMENTS

Special thanks to the workers and students at the Brayton Firefighter Training Field, especially Chief Cary Rocaforte. This material is based in part upon work supported by the National Science Foundation under Grant Numbers IIS-0803854 and IIS-0742947. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Alexander, A.L., Brunyé, T., Sidman, J., Weil, S.A. From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-In and Their

- Effects on Transfer in PC-Based Simulations and Games. *DARWARS Training Impact Group Whitepaper* (Nov 2005).
- [2] Allen, J. Maintenance Training Simulator Fidelity and Individual Differences in Transfer of Training. *Human Factors* 28 (1986), 497–509.
 - [3] Beaubien, J.M., Parker, D.P. The Use of Simulation for Training Teamwork Skills in Health Care: How Low Can You Go? *Quality and Safety in Health Care* 13 (2004), 51–56.
 - [4] Deneff, S., Ramirez, L., Dyrks, T., Stevens, G. Handy Navigation in Ever-Changing Spaces: An Ethnographic Study of Firefighting Practices. *Proc. ACM Designing Interactive Systems* (2008), 184–192.
 - [5] Entin, E.E., Serfaty, D. Adaptive Team Coordination. *Human Factors* 41, 2 (June 1999), 312–325.
 - [6] Gentner, D., Stevens, A.L. *Mental Models*. Lawrence Erlbaum Assocs., Hillsdale, NJ, USA, 1983.
 - [7] Hays, R.T., Singer, M. J. *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training*. Recent Research in Psychology. Springer-Verlag, New York, NY, USA, 1989.
 - [8] Hutchins, E. *Cognition in the Wild*. MIT Press, Cambridge, MA, USA, 1995.
 - [9] Kozlowski, S.W.J., DeShon, R.P. A Psychological Fidelity Approach to Simulation-Based Training: Theory, Research and Principles. *Scaled Worlds: Development, Validation, and Applications*, S. G. Schiflett, L.R. Elliott, E. Salas, M.D. Coover, Eds. Ashgate Pub. Ltd., 2004, 75–99.
 - [10] Mathieu, J.E., Goodwin, G.F., Heffner, T.S., Salas, E., Cannon-Bowers, J.A. The Influence of Shared Mental Models on Team Process and Performance. *Journal of Applied Psychology* 85, 2 (2000), 273–283.
 - [11] MacMillan, J., Entin, E.E., Serfaty, D. Communication Overhead: The Hidden Cost of Team Cognition. *Team Cognition: Understanding the Factors that Drive Process and Performance*, E. Salas, S.M. Fiore, Eds., 1st ed. American Psychological Association, Washington, DC, USA, 2004, 61–82.
 - [12] Reynolds, C.W. Flocks, Herds and Schools: A Distributed Behavioral Model. *Proc. ACM SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques* (1987), 25–34.
 - [13] Rehmann, A.J., Mitman, R.D., Reynolds, M.C. A Handbook of Flight Simulation Fidelity Requirements for Human Factors Research. Technical Report DOT/FAA/CT-TN95/46, Crew System Ergonomics Information Analysis Center, Wright-Patterson AFB, OH, Dec 1995.
 - [14] Salen, K., and Zimmerman, E. *Rules of Play: Game Design Fundamentals*. MIT Press, Cambridge, MA, USA, 2004.
 - [15] Simon, H.A. *Sciences of the Artificial*, 3rd ed. MIT Press, Cambridge, MA, USA, 1996.
 - [16] St. Julien, T.U., Shaw, C.D. Firefighter Command Training Virtual Environment. *Proc. Conference on Diversity in Computing* (2003), 30–33.
 - [17] Suchman, L. A. *Plans and Situated Actions: The Problem of Human Machine Communication*. Cambridge University Press, Cambridge, UK, 1987.
 - [18] Tate, D.L., Sibert, L., King, T. Using Virtual Environments to Train Firefighters. *Computer Graphics and Applications* 17, 6 (Nov-Dec 1997), 23–29.
 - [19] Toups, Z.O., Kerne, A., Hamilton, W.A., Shahzad, N. Zero-Fidelity Simulation of Fire Emergency Response: Improving Team Coordination Learning. *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems* (2011), 1959–1968.
 - [20] Toups, Z.O., Kerne, A., Hamilton, W.A. The Team Coordination Game: A Zero-Fidelity Simulation from Fire Emergency Response Practice. *ACM Trans. Computer-Human Interaction* (2011), in press.
 - [21] Toups, Z.O., Kerne, A. Implicit Coordination in Firefighting Practice: Design Implications for Teaching Fire Emergency Responders. *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems* (2007), 707–716.

5.13 Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions

Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions

Michael W. Martin
Old Dominion University
mmart081@odu.edu

Yuzhong Shen
Old Dominion University
yshen@odu.edu

Abstract. This paper explores the distinction between operative and resultant actions in games, and proposes that the learning space created by a serious game is a function of these actions. Further, it suggests a possible relationship between these actions and the forms of cognitive load imposed upon the game player. Association of specific types of cognitive load with respective forms of actions in game mechanics also presents some heuristics for integrating learning content into serious games. Research indicates that different balances of these types of actions are more suitable for novice or experienced learners. By examining these relationships, we can develop a few basic principles of game design which have an increased potential to promote positive learning outcomes.

1.0 INTRODUCTION

In his book, *What Video Games Have to Teach Us About Learning and Literacy*, J. P. Gee alludes to the concept of a learning space while describing the creation of an educational *psychosocial moratorium* within serious games:

"...I learned that video games create what the psychologist Eric Erikson has called a psychosocial moratorium – that is, a learning space in which the learner can take risks where real-world consequences are lowered [1]."

While Gee does not expound further on the concept, the phrase "learning space" may be inferred to mean the world created by the video game. When a learner engages with the game world, they do so with few of the risks that might be inherent in performing similar activities in the real world.

For the purposes of closely examining or creating a learning space, it can be helpful to examine this concept further. In doing so, we might discriminate more strictly between the game environment and the learning space. A learning space is likely to not include

everything that constitutes the game. The learning content may be restricted to only certain subset of the mental space defined by the game, with other portions of the game having little or no bearing on the intended instruction. Based on this paradigm, a serious game designer should consider how the distinction between learning space and non-learning space impacts learning outcomes, and accordingly, where the learning content belongs in the game.

Insight into these questions can be found in an examination of player actions within games. J. Schell describes a design model which distinguishes between the types of actions that a player has available to them. This categorization of actions can be used to characterizing the depth, complexity and elegance of the game. By moderating the ratios of action types, the game designer can create distinct experiences for the player. Similarly, an instructional designer might engage in the same sort of moderation, not only in developing player actions, but additionally in deciding which types of actions will reflect the learning content.

The combinations of actions and learning content also bear on the cognitive load

induced upon the learner. Research by Park, et al, and Lee et al, demonstrates a relationship between interaction levels, learner proficiency levels, and learning outcomes [2, 3]. By carefully constructing the actions available to the user, a designer has a set of tools through which it may be possible to create serious games that achieve desired pedagogical outcomes.

2.0 DISCUSSION

2.1 Games, Learning Spaces, and Learning Content

Simple observation seems to indicate that in most cases, the learning content in serious games does not make up the entirety of the experience. Games are likely to include aspects which are not related to the instructional material. This can be seen most readily when the motif of a game, as defined by C. Totten [4], is not specifically dictated by or dependent upon the learning content. Totten distinguishes between the mechanics and the motif of a game, with motif comprising aspects of the game like narrative and themes. Motif is the aesthetic or thematic presentation of the game beyond the mechanics which embed the rules. Hunicke, et al. similarly associate game aesthetics with the emotional “fun” response evoked by the player interacting with the game [5].

In the business model simulation game *Lemonade Stand*, shown in Fig. 1, the distinction between aspects of the motif and the learning content are clearly visible [6]. The game is intended to teach about the challenges of running a business, but the aesthetics of dogs, cats, and strange creatures have no bearing upon that content.

Even while serious game developers and researchers stress the importance of integration of game fantasy and learning content, there is still evidence that the motif

and mechanics do not entirely overlap with the learning content.

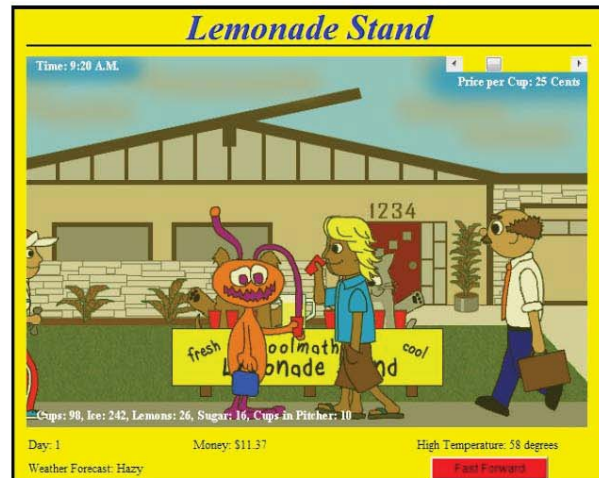


Fig. 1. Lemonade Stand business simulation game depicting the distinction between the game and the learning content [6]. The learning content is focused on the challenges of purchasing good and making a profit, and have nothing to do with the animals and strange creatures shown in the screenshot.

M. Habgood, et al. created a game to teach division skills and demonstrated the importance of endogenous fantasy in which the fantasy of the game is intrinsic to the learning content and vice versa. Yet a screen shot of the game still shows that it is challenging to not include aesthetic content in a game that is not necessarily related to the learning content [8]. Figure 2 shows Habgood's *Zombie Division* game, in which the learner must use numerically oriented weapons to “divide” the enemy skeletons.

Even though the game is cited as an example of intrinsic fantasy, it is clear to see that the motif is not directly relevant to mathematical division. The aesthetics of the game depict a small Greek warrior attacking skeletons, and the linkage between these fantasy characters and division is created through the mechanics of the game, not the motif.



Fig. 2. Screenshot of *Zombie Division*, in which the learner controls a Greek Warrior and uses the principles of division to defeat skeleton opponents [7]. As with *Lemonade Stand*, the aesthetics or motif of the game have almost no connection to the learning content.

Based on these examples, it may be concluded that the learning space is a subset of the larger game environment, and that further, the learning space is a result of the interaction of learning content with the game elements (mechanics and motif). In part, the learning content may be presented in the game or simulation through expository methods, such as narratives, cut scenes, or non-interactive demonstrations. But the learning space which the game users actively explore is created when the learning content is made available to the user through game-play. It is through these mechanics that the user has a chance to meaningfully interact with the content. Figure 3 depicts the relationship between Game, Learning Content, and Learning Space. The next challenge is to understand how designers create game-play for the users.

2.2 Operative and Resultant Actions

In *The Art of Game Design*, Schell discusses two important concepts which play a significant role in shaping game mechanics. These concepts are the users operative and

resultant actions [9].

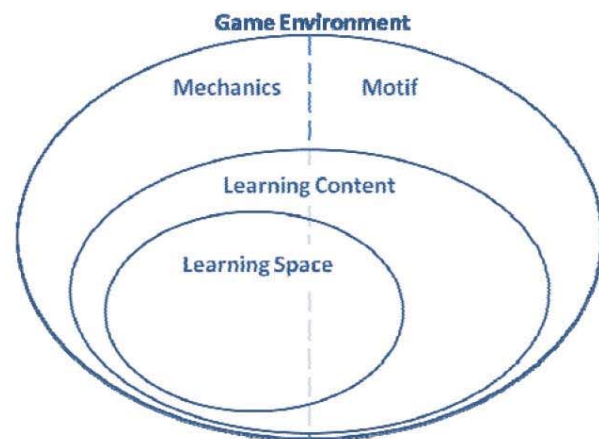


Fig. 3. The relationship between Game Environment, Learning Content, and Learning Space. It is likely that Learning Space derives primarily from Mechanics side of the Game Environment, but it is possible to derive portions of it, as well, from the Motif.

Operative actions are the basic actions which a player might engage in. In a game of checkers, these operative actions might include moving a checker forward, jumping an opponent, or moving a king backwards. Due to the nature of computers, video games must explicitly delineate these actions and their effects. As a result, in computer games, these actions tend to be very discrete and well defined.

Resultant actions, on the other hand, are the “meta” actions which the player can take in order to achieve a goal. In the example of checkers, the player might force an opponent to make an unwanted jump, or protect a piece from being taken. Both of these resultant actions might take the form of the same operative action (moving a checker forward), but they serve different purposes, and are enacted for different reasons. Schell describes how these resultant actions are often not concrete parts of the game, but rather are aspects which develop through game play. As such, they are ill-defined, and more subjective than operative actions. To an

extent, these resultant actions are analogous to strategies developed to achieve goals within the game. Even in games implemented on video games, which tend to require stricter definition due to the nature of computer programs, these actions have a larger degree of latitude than their operative counterparts.

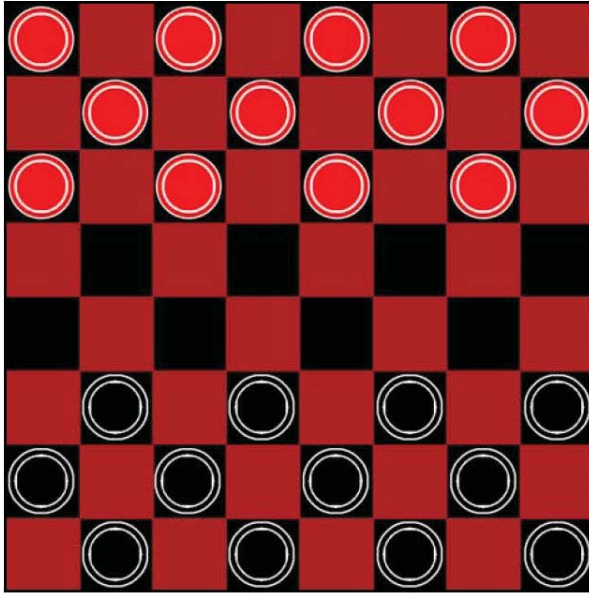


Fig. 4. Checkers Board. Each piece only provides the player with a few operative actions, but the game still has many resultant actions.

Schell also discusses the interaction of operative and resultant actions, and their effect upon the complexity and the potential for emergence in a game. From this discussion, it is possible to discern some basic formulae with which to characterize the games in terms of complexity, elegance and depth. The following are proposed definitions of these terms, in the context of Schell's actions.

Complexity may be thought of as a reflection of the operative actions. Games are complex when the user has many specified actions they can take. A game of chess, for example, is a relatively more complex game than checkers. The chess player has many operative actions; each side consists of

sixteen pieces, made up of six unique types. These unique types each behave differently, providing the player with distinct operative actions for each piece on the chess board. In contrast, as mentioned above, checkers has fewer operative actions for the player to consider.

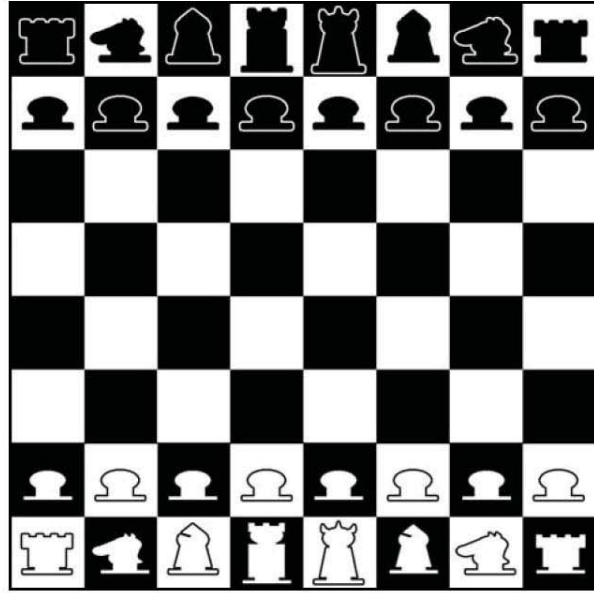


Fig. 5. Chess Board. Each type of piece has unique rules that govern its movement, effectively giving the player a larger number of operative actions than Checkers.

Elegance, on the other hand, can be considered the ratio of resultant actions to operative actions. Games that are elegant have more resultant actions in comparison to the number of operative actions. The game of Go is a good illustration of elegance. Go only has two operative actions – to place a stone on the board, or to pass. All stones are identical, and behave the same, and the action of placing the stones is the same every time. However, Go, like Chess, still has an abundance of resultant actions.

That abundance of resultant actions sums up the last characterization: depth. Schell suggests that one way to create the potential for resultant actions through the addition of clever and interactive operative actions.

However, he simultaneously cautions that too many poorly considered operative actions can result in a game that is “bloated, confusing, and inelegant” [9]. Games which are elegant achieve depth while keeping complexity to a minimum.

From the Instructional Design perspective, Morrison, et al. provide specific prescriptions for designing instruction based on the type of learning content [10]. In examining these prescriptions, it seems likely that it is in the depth of the game, rather than the complexity of the game, that the user might engage in activities that would support teaching principles, rules, and procedures. In contrast, if the learning content is limited to facts and concepts, requiring lower levels of Bloom’s Taxonomy [11], then it may be sufficient to embed the learning content into the operative actions, or even in the game motif.

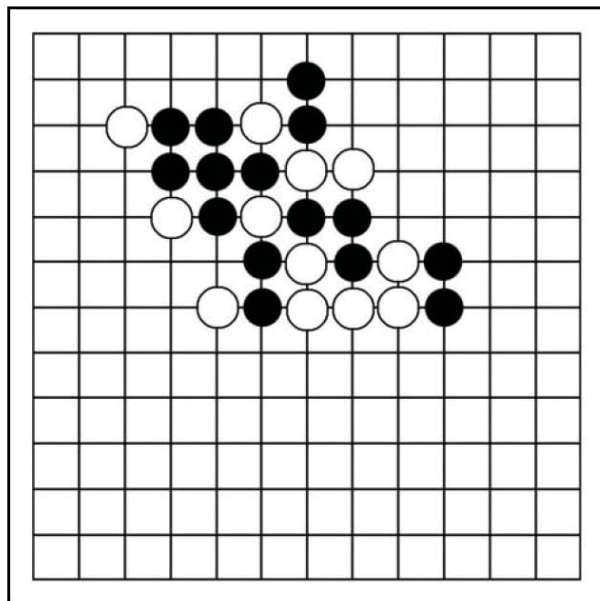


Fig. 6. Go board. With only two operative actions, the game of Go still manages to create a very deep game play experience, embodying the notion of elegance.

Schell’s operative and resultant actions provide the serious game designer with a means by which to instill desired levels of complexity and depth into a game. By

manipulating these elements of the game design, game developers can attune their game design to support instructional needs. But the potential pedagogical implications for operative and resultant actions are even farther reaching.

2.3 Actions and Learner experience Levels

The balance between operative and resultant actions may have bearing not only on the learning space created and the elegance of the game, but may also on the type of learner for which the game is designed.

S. Park, et al. studied the use of simulations with high and low levels of interactivity by experienced and novice learners [2]. One conclusion drawn was that experienced learners did better with more complex simulations, while inexperienced learners performed relatively worse. Within the Park’s research, the difference in the complexity of the simulations can be largely described as a difference in the amount of operative actions given to the users. The results of their research also indicate that the mean cognitive load score for inexperienced students increased along with the increase in complexity, while the mean score for experienced students decreased. Figure 7 shows the results from Park, et al.’s research.

Based on this research, it would follow that advanced learners would benefit more for having many operative actions through which to interact with their serious games. Designers targeting such learners would be free to include many operative actions, presumably creating both complexity and depth. In contrast, serious games designed for novice learners should focus on elegance, reducing the number of operative actions in order to reduce the complexity of the game interactions.

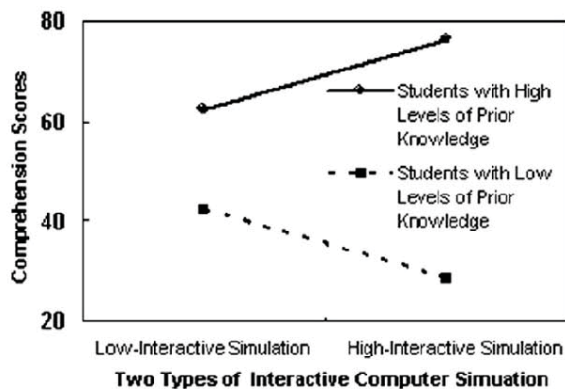


Fig. 7. Results of Experiments conducted by Park, et al. Students with high levels of prior knowledge performed better with high-interaction (and therefore higher complexity) simulations than those with low levels of prior knowledge.

2.4 Learning Curves

The results of Park, et al.'s research bear a remarkable resemblance to the hypothetical learning curve suggested by S. Alessi [12]. Figure 8 depicts Alessi's curve. Park's graph only presents 4 data points, making it difficult to ascertain whether the data fully describes the curve, but experimental results seem to support the hypothetical curve.

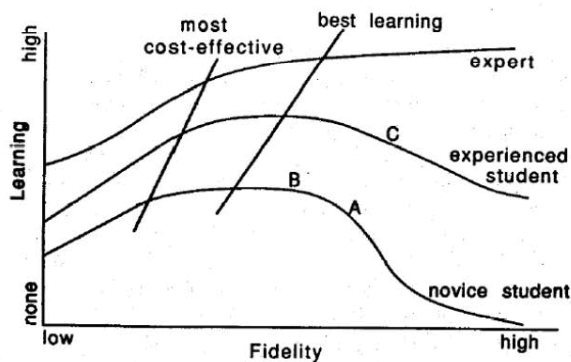


Fig. 8. Alessi's Hypothesized Learning Curve. Alessi proposes that more experienced students will benefit from higher fidelity simulations, while the performance of novice students will decline after reaching a certain optimal level.

Alessi's curve is hypothetical, and his article concludes by stating that future research should assess what aspects of a simulation should be varied and under what conditions this variation should occur. Park's data suggests that interactivity is one place in which variation should occur, and Schell's actions further suggest that manipulating the operative actions made available to a learner may be a good prescription for calibrating a game to the needs of the learner.

It should be noted that Alessi is comparing learning to simulation fidelity, while Park, et. al. are describing learning (as reflected by comprehension scores) against the levels of interactivity in the simulation. In order to make the leap between the two graphs, it must be presumed that interactivity is correlated with simulation fidelity.

2.5 Cognitive load

The justification for making this leap between interactivity and simulation fidelity may be found in J. Sweller's definition of cognitive load, and it's three components: intrinsic, extraneous, and germane load [13].

Intrinsic cognitive load results from the difficulty of the material being dealt with. Extrinsic cognitive load is created by the manner in which the material is presented, and lastly, germane cognitive load is the load associated with processing and encoding schemas. Alessi defined fidelity as the degree to which a simulation reflects reality, or the phenomena being presented in the instructional content [12]. Park's presentation of simulation complexity via interactivity is a clear reflection of the intrinsic load being imposed on the learner [2]. Based on these definitions, simulation fidelity correlates to the intrinsic cognitive load, while simulation or game interactivity correlates to extrinsic cognitive load.

Complexity and fidelity form two sides of the cognitive load triumvirate, and experimentation by Lee, et al. concludes that adjustments of presentation (and

corresponding intrinsic load) can be used to calibrate the overall cognitive load [3]. Additionally, their experimental results are similar to those of Park, et al., though their experimental design included an additional factor comparing iconic versus symbolic representation. Increasing or decreasing the extraneous cognitive load presented to the learner can compensate for their experience level with the learning content. If the learner is experienced, then they can be effective with complex presentations, whereas novice learners should be given more simple interfaces in order to account for their position on the learning curve with respect to the intrinsic difficulty of the subject matter. As before, the moderation of the extraneous cognitive load can be performed by adding or removing operative actions.

3.0 CONCLUSION

This paper has presented a discussion of the nature of conceptual learning spaces within in serious games. By examining selected research and theoretic constructs, it may be possible to enact specific game design principles to elicit desired learning outcomes. Specifically, this paper proposes that:

- Learning spaces are not synonymous with game environments, but rather are a subset of learning content, which in turn, is a subset of the game environment.
- Learning spaces are functions of the interaction of the learning content with the game mechanics and with the game motif.
- Fact and concept categories of learning content may be taught through learning spaces based primarily on operative actions and motif within games.
- Rule, principle and procedure categories of learning content may be taught through learning spaces based primarily on resultant actions within games.
- Novice learners may be better served with

lower complexity games with fewer operative actions.

- Learners of increasing experience may benefit more from added operative actions.

While these conclusions are based on existing research and experimentation, there is a need to confirm these propositions with additional causal studies. Much of what is proposed here is based on small leaps of knowledge that could benefit greatly from relevant empirical data.

REFERENCES

- [1] J. P. Gee, *What Video Games Have to Teach Us About Learning and Literacy*, Rev. and updated ed. New York: Palgrave Macmillan, 2007.
- [2] S. I. Park, et al., "Do Students Benefit Equally from Interactive Computer Simulations Regardless of Prior Knowledge Levels?," *Computers & Education*, vol. 52, pp. 649-655, 2009.
- [3] H. Lee, et al., "Optimizing Cognitive Load for Learning From Computer-Based Science Simulations," *Journal of Educational Psychology*, vol. 98, p. 902, 2006.
- [4] C. Totten, "Mechanics vs. Motif," in *Christopher Totten's Blog* vol. 2010, ed: UMB TechWeb, 2010.
- [5] R. Hunicke, et al., "MDA: A Formal Approach to Game Design and Game Research," presented at the Challenges in Games AI Workshop, Nineteenth National Conference of Artificial Intelligence, San Jose, California, USA, 2004.
- [6] D. Ramey. (2010, 7 October). *Coolmath's Lemonade Stand Game*. Available: <http://www.coolmath-games.com/lemonade/>
- [7] M. P. J. Habgood, "The Effective Integration of Digital Games and Learning Content," Philosophy, University of Nottingham, 2007.

- [8] M. P. J. Habgood, *et al.*, "Endogenous Fantasy and Learning in Digital Games," *Simulation & Gaming*, vol. 36, pp. 483-498, 2005.
- [9] J. Schell, *The Art of Game Design : A Book of Lenses*. Amsterdam ; Boston: Elsevier/Morgan Kaufmann, 2008.
- [10] G. R. Morrison, *et al.*, *Designing effective instruction*, 5th ed. Hoboken, NJ: J. Wiley, 2007.
- [11] L. W. Anderson and L. A. Sosniak, *Bloom's Taxonomy : A Forty-Year Retrospective*. Chicago: NSSE : Distributed by the University of Chicago Press, 1994.
- [12] S. Alessi, "Fidelity in the Design of Instructional Simulations," *Journal of Computer-Based Instruction*, vol. 15, pp. 40-47, 1988.
- [13] J. Sweller, "Cognitive Load Theory, Learning Difficulty, and Instructional Design," *Learning & Instruction*, vol. 4, pp. 295-312, 1994.

5.14 Innovating Training through Immersive Environments: Generation Y, Exploratory Learning, and Serious Games

Innovating Training through Immersive Environments: Generation Y, Exploratory Learning, and Serious Games

Gerald "Jay" Gendron
gerald.gendron@gmail.com

Abstract. Over the next decade, those entering Service and Joint Staff positions within the military will come from a different generation than the current leadership. They will come from Generation Y and have differing preferences for learning. Immersive learning environments like serious games and virtual world initiatives can complement traditional training methods to provide a better overall training program for staffs. Generation Y members desire learning methods which are relevant and interactive, regardless of whether they are delivered over the internet or in person. This paper focuses on a project undertaken to assess alternative training methods to teach special operations staffs. It provides a summary of the needs analysis used to consider alternatives and to better posture the Department of Defense for future training development.

1.0 INTRODUCTION

There are growing challenges in military training – specifically the staff training of special operations forces. This paper highlights a recent project to assess learning approaches used for staff training. The project primarily focused on conducting a needs analysis and implementing a new method based on the information learned. The new method is a type of serious game which was delivered to a military unit in the Pacific at the end of the project. This serious game not only satisfied an existing need, but it is shaping the culture of those who develop training such that virtual worlds and other immersive environments may be welcomed and used in the coming years.

In addition to providing a review of the project, this paper will also pose questions to consider about the near- and long-term. The needs analysis which enabled the implementation of the serious game also identified some questions that remained unaddressed at the end of the project timeline. The project team continues to assess the questions through a requirements generation process to determine how follow-on projects could better address the challenges identified.

This paper discusses the project for a special operations unit, but the lessons and

insights could be applicable to situations in other governmental agencies, learning institutions, and corporate training environments.

2.0 SITUATION

2.1 Problem Statement

The United States Armed Forces face many challenges in this time of constrained resources. The Obama Administration has established a goal to reduce Defense Department budgets by approximately \$400 billion through 2023 [1]. Meanwhile, there is a growing requirement in joint military training to

Improve staff training at the operational level of war among teams distributed around the globe – both inside and outside of the military establishment.

This *tyranny of distance*, as the time and space aspects of training have been called, is challenging. In addition to that, a changing demographic among middle-grade staff officers could make the overall situation more challenging than expected. Officers entering the grades of major and lieutenant commander will often start with a staff tour at one of their Service headquarters or with a joint organization like

the Joint Staff or a combatant command. These arriving staff officers have historically posed a challenge because they must quickly acclimate and perform in a significantly different environment with unfamiliar processes. Training developers have addressed these needs in the past with annual, live training events.

A study [2] commissioned in 2008 by the Joint Staff J7's Joint Training and Exercise Division found officers serving on operational-level joint staffs were not performing as well as senior leaders had expected. Reasons cited by the study include the facts that these officers were on their first joint tour, their first staff tour, and their first time working with a diverse workforce such as other military Services and other agencies. Experience points to a lack of process-oriented training on staff techniques and skills. Newly arriving staff members must seek on-the-job training. As a training customer, they represent an underserved audience during their first four to eight months of time on the staff.

Within the next few years, the new staff officers will come from Generation Y. This new demographic becomes an additional factor for training developers to contend with while assessing new approaches. The question becomes

As Generation Y joins military staffs, what are the challenges and insights training developers should recognize?

This problem also applies to special operations staffs. To address the problem, the project team's initial concept was to increase the number of training events between large-scale, annual exercise. It was suggested that this could sustain performance levels. However, any solutions considered could not result in additional travel time or costs and it had to facilitate staff training of complex scenarios.

2.2 Situational Context

Special operations forces represent a unique arm of national security. Many people are familiar with special operations from the dangerous and heroic missions that have recently made the news. Less known – but of great importance – are the diplomatic aspects of special operations. When not in a combat zone, these mature and highly experienced operators are more often engaged in supporting the United States' diplomatic mission through military-to-military training with partners and allies. They also support the Department of State in training and educating people from the United States Government or allied nations in military planning and execution processes. These types of missions are primarily focused on *building and maintaining relationships*. Performing this mission over long distances is an added challenge.



Figure 1. Distributed nature of military units.

Consider the map shown in Figure 1. It shows the locations of a Theater Special Operations Command (TSOC) Headquarters in the Pacific along with their assigned components in Japan and Guam. This command is spread across 4,000 miles, yet they must be able to operate as well as any other staff which resides in the same location. Effective military training helps make this possible.

TSOCs are operational level staffs that focus on planning and executing special operations missions for the combatant command to which they are assigned. They maintain a regional focus whether that be the Pacific, Europe, or the Middle East. There is also a unique TSOC that focuses on training. Special Operations Command Joint Capabilities is charged with the mission to train TSOC staffs in planning processes. They have traditionally done this through academic sessions in a live training environment and through computer-based training.

A word about the military men and women who work on joint staffs: A significant level of a staff's membership is comprised of majors and lieutenant commanders. Around the hallways of the Pentagon they are called *Iron Majors*. This is a term of endearment as they are the heart and soul of a staff – the action officer. Although there are officers senior to them and a strong enlisted force supporting them, the Iron Majors are the workhorse of a joint staff. Why is this? Majors and lieutenant commanders have attained a level that requires them to learn a new skill set. Prior to their arrival on the staff – for about 11 years – they have trained and executed tactical level missions and have become proficient within their own Services. Before moving on to command positions as lieutenant colonels and naval commanders, these Iron Majors are refined in the crucible of a staff where they are asked to take their experiences to new levels. This same situation occurs at the TSOCs. Although TSOC staffs are much smaller than Service and Joint Staffs, the same dynamic is at play.

3.0 NEEDS ANALYSIS

Training of special operations staff officers has essentially the same objective as other training – transmitting knowledge and imparting learning to a group using some

type of medium. Good training and education products are based on good pedagogy, and they are informed by a solid needs analysis of the training audience. This is the underpinning of the Instructional System Design method [3]. In the case of the special operations training project, a team of developers decided to conduct a needs analysis to characterize requirements for next generation staff training. They approached the problem by establishing a framework of three primary questions to consider alternatives while ensuring critical elements were considered:

- Who is Generation Y which will soon be a primary training audience?
- What is known about their preferences for learning?
- How might training methods need to change relative to the audience?

3.1 Generation Y and Exploratory Learning

In 2011, the majority of the Department of Defense and TSOC workforce falls into one of three generations as defined by sociologists. Figure 2 shows the generational groups: Baby Boomers (1946-1964), Generation X (1965-1980), and Generation Y (1981-2000). As a way of highlighting the differing experiences of each generation, the figure shows just two of the many technological changes the



Figure 2. Generations in the workplace, birth years, and technological change.

generations have experienced over time. In fact, members of Generation Y are currently in the military and they have bravely served in the battlefields of Iraq and Afghanistan. Generation Y will begin to arrive in staff positions around the year 2014, if not a bit sooner. This includes their arrival in TSOCs around the globe.

The Joint Staff study [2] identifies problems in training programs for newly arriving staff officers. Note that it was conducted in 2008, before Generation Y was a part of the study population. Experience from the battlefields has shown that Generation Y thinks and acts differently than earlier generations. Training developers at Special Operations Command Joint Capabilities are officers and support contractors experienced in the areas they train to TSOC staffs, but they are not from Generation Y. The resulting situation: those responsible for training development and making decisions on the next generation of training technologies may have vastly different ideas of what is needed by the training audience.

There is a body of work in the literature about Generation Y and their similarities and differences as compared to other generations. As early as 1996, Rieman, Young, and Howes [4] had already identified differences in learning styles among the newest generation. They found Generation Y preferred Exploratory Learning which they characterized by four elements:

- Task oriented
- Time constrained
- Primary goal is *task performance*
- *Learning* is a secondary aspect

How is Generation Y performing differently and does this suggest a need for different training? Yes. While Generation Y has not entered the staffs yet, it was noted they have been serving in our military. Reports from trainers who have worked in Iraq and Afghanistan indicate the junior officers

(Generation Y) have a different approach to operations and learning. In fact, the Department of Defense has commissioned experiments which have specifically studied various elements of junior officers ranging from the need for new training to suggestions that senior leaders increase their trust in the junior officers.

The author wishes to share the experiences of Mr. John Hunter, a fourth grade school teacher, who was invited to present at TED2011 in Long Beach California [5]. John's story is instructive of how a different generation learns – demonstrating how learning and the process of training development may be approached from multiple perspectives and should be a creative undertaking. John makes a key statement:

My question to you is, who's in charge of that classroom? It's a serious question: who is really in charge? I've learned to cede control of the classroom over to the students over time. There's a trust and an understanding and a dedication to an ideal that I simply don't have to do what I thought I had to do as a beginning teacher: control every conversation and response in the classroom. It's impossible. Their collective wisdom is much greater than mine, and I admit it to them openly [5].

When it comes to learning – Generation Y appears to be different. Collaboration through a designed task is important. It has been documented by scholars, observed on battlefields, and acknowledged in classrooms. Trainers and training developers must address this aspect of learning in their approaches.

3.2 Serious Games

The analysis thus far has indicated the need to focus on a changing demographic in the training audience ("who"). It also found literature and experience that indicates a difference in learning style ("what") among this new demographic. This led the project team to consider "how" TSOC training methods need to change relative to the audience preferences.

There are two basic categories of TSOC training: traditional training and immersive learning environments. Traditional training encompasses a host of training tools and methods ranging from computer-based training to live training and written material. TSOC members have the means to take an introductory training course using computer-based lessons. In fact, they are prerequisite for attendance to more advanced, live TSOC training events. Additionally, reports promulgating best practices are published for individual study. A pro of traditional training is its familiarity among trainers and training audiences. Most people in the joint community have been working as professionals in their Services for a decade or more. During this time they have likely come into contact with training exercises and grown accustomed to the current training methods. There is ample history which indicates traditional training is effective.

However, there are cons related to traditional training. Budget constraints and demanding operations tempo among trainers and the training audience restrict live training offerings to a small number each year. In most cases, this translates into a single training event. It is arguable that the indirect costs are substantial. These costs range from poor readiness among the workforce to an unintended inertia hindering new techniques. While computer-based training is more sophisticated than written material, it does

not instill a high degree of interaction among users. According to Windham [6], live classroom training is valued because it satisfies a desire for interaction. This craving for interaction will persist into online environments. For this reason, computer-based training's lack of interaction places it in a class similar to written materials.

Immersive learning environments (ILEs) comprise the other alternative looked at in this project. ILEs are "learning situations that are constructed using a variety of techniques and software tools including game-based learning, simulation-based learning and virtual worlds" ([7], p. 2). ILEs can now be found in government organizations. Examples include the CyberProtect game developed by the Defense Information Systems Agency and virtual worlds developed by Defense Acquisition University. These games or worlds harness simulation-based learning to drive a scenario for the audience. Many pros of ILEs can be found in the literature. McNeely [8] notes students often learn things best by doing. Gee [9] argues that what people "are doing when they are playing video games is often good learning" (p. 199). There are, however, cons associated with ILEs. They are characteristically more costly and sophisticated than traditional training methods. This requires a more thorough needs analysis to ensure developing an ILE provides a better training experience than some other method [7]. Nonetheless, the author believes ILEs like serious games will increasingly be used for future TSOC training.

In the TSOC project, a serious game-based tool was used to enhance traditional training approaches. A feasibility assessment was conducted to look at game functionality and user acceptance [10]. Training audience surveys indicated a number of positive attributes of the serious game approach. Highlights of that assessment indicate

serious games are best accepted by TSOC staffs when they:

- focus on key TSOC procedures,
- allow for distributed players, and
- support tailored scenarios.

The assessment found the serious gaming approach was capable of allowing staffs to rehearse necessary planning functions from distributed locations without traveling to a central location. The essential point here is staffs perform best when key procedures are practiced often. These procedures and checklists are what drive staffs during crisis situations. Secondly, not only do staffs desire additional training, but they also desire access to skilled observers who can provide feedback. The web-based, distributed approach of the game provided a capability for trainers from Special Operations Command Joint Capabilities to reach out to training audiences from their home location. Although Special Operations Command Joint Capabilities cannot physically travel to a TSOC more than one or two times per year, a distributed game allows them to support the TSOC requests. Finally, the serious game approach delivers special operations staff training using a tailored scenario at a time and place of the audience's choosing [10]. Staffs do not value "universal" training scenarios. Special operations staffs work at a regional focus and any training must support tailored scenarios in a cost effective process.

4.0 A LOOK TOWARDS THE FUTURE

We now accept the fact that learning is a lifelong process of keeping abreast of change. And the most pressing task is to teach people how to learn. —Peter Drucker

Observations made during the TSOC project helped to successfully deliver a serious game-based training method to a

TSOC in 2011. The project sought to determine the needs of a changing staff demographic, compare those needs to existing training approaches, and introduce new approaches when gaps were identified. The project found enough evidence to suggest Generation Y has different training needs. It also found that immersive learning environments like serious games can satisfy those needs.

In the future, the concept for TSOC serious games is for them to fully complement annual live training. This is expected to maintain proficiency levels closer to optimal levels – thereby reducing the performance degradation currently witnessed in once-per-year training cycles. This new, responsive training method uses distributed and virtualized training technologies to establish simulated scenarios. It can be delivered to military as well as other agencies and partners. Training becomes not only more frequent but also more efficacious in developing our human capital.

The serious game used in this project appears to be a leverage point to prepare for the use of virtual worlds in staff training. In essence, they provide an evolutionary approach. The introduction of serious games in special operations staff training is helping to shift the culture among training audiences and training developers alike – this shift has been witnessed in recent briefings presented by Special Operations Command Joint Capabilities. It is this author's opinion that greater adoption of virtual environments will provide important capabilities like observing a training audience and their "physical" activities. These activities are a very important aspect of live training which can be observed in traditional, live training – such as observing to see if a player moves to another area of a joint operations center to gather information and collaborate with others. Additionally, the game-like feel of the scenario instills a sense of "play" and interest into the training

event. The Special Operations Command Joint Capabilities experience marks a beginning of a transition towards virtual worlds for staff training.

Finally, the project successes provide interesting opportunities to think about the broader future of staff training. It is important to consider what the future holds and develop questions to make sure the right training and learning is delivered to the desired audience with an appropriate leveraging of technology.

What technologies will be available in future joint training events?

What is the role of technology in enabling new processes and relationships?

To summarize, this project focused on assessing needs of Generation Y within TSOC staffs and using that analysis to explore immersive learning environments. Three overarching observations are captured here as insights to future training design:

- Processes used to train staffs must adapt to the influx of Generation Y arriving for their first staff assignments.
- Technology adoption strategies must also take into account needs analysis of Generation Y.
- Training strategies used to deliver content must consider new technologies identified in needs analyses.

Over the next decade, young men and women who have served in the combat zones of Iraq and Afghanistan will be coming home to serve their tours at Service and joint staffs. They come from a different generation than those currently in leadership and decision making positions. Immersive learning environments include

serious games and virtual worlds. Generation Y has certain demands and expectations of learning environments. This paper affords some thoughts and questions for consideration about how future training can best serve Generation Y by adapting learning methods while leveraging a proper balance of technology to deliver the learning.

5.0 REFERENCES

- [1] Winter, D.C. (2011, May 31). *Defense planning for the twenty-first century: Lessons from the twentieth* (Report of the Heritage Foundation). Retrieved from <http://www.heritage.org/research/reports/2011/05/defense-planning-for-the-twenty-first-century-lessons-from-the-twentieth>.
- [2] United States Special Operations Command. (2010, October). *US special operations command joint staff officer site visit*. Tampa, FL.
- [3] Carliner, S. (2008). A holistic framework of instructional design for e-learning. In Carliner, S. & Shank, P. (Eds.), *The e-learning handbook: Past promises, present challenges* (pp. 307-358). San Francisco, CA: Pfeiffer.
- [4] Rieman, J., Young, R. M., & Howes, A. (1996, June). A dual-space model of iteratively deepening exploratory learning. *International Journal of Human-Computer Studies*, 44(6), 743-775.
- [5] Hunter, J. (Presenter). (2011, March). The World Peace Game. *TED talks*. Podcast retrieved from http://www.ted.com/talks/lang/eng/john_hunter_on_the_world_peace_game.html.

- [6] Windham, C. (2005). The student's perspective. In D. G. Oblinger & J. L. Oblinger (Eds.), *Educating the net generation*. (pp. 5.1 – 5.16). Washington, DC: Educause.
- [7] Rozwell, C. (2008, October 31). *Learning simulations equip people to face crises* (Gartner Report G00162391). Stamford, CT: Gartner, Inc.
- [8] McNeely, B. (2005). Using technology as a learning tool, not just the cool new thing. In D. G. Oblinger & J. L. Oblinger (Eds.), *Educating the net generation*. (pp. 4.1 – 4.10). Washington, DC: Educause.
- [9] Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Available from <http://site.ebrary.com/lib/nationaldefense/Doc?id=10089175&ppg=205>.
- [10] Special Operations Command Joint Forces Command. (2010). *Quick look report for distributed IW training: Small group scenario trainer*. Suffolk, VA.

5.15 The Role of Game Based Learning in the Health Literacy of African American Adolescent Males

The Role of Game Based Learning in the Health Literacy of African American Adolescent Males

Judith Connell and Margaret Knight
Norfolk State University
jconnell@nsu.edu mdknight@nsu.edu

Abstract. Twenty-first century literacy is more than being able to encode for spelling ability, decode for reading comprehension, and calculate for numeric reasoning. It demands the skills to negotiate the world of technology. Health literacy is lower than general literacy, and general literacy is lower among African American males than the overall population. The authors discuss the prospects of incorporating Game Based Learning approaches into strategies for teaching health literacy. Results of a survey administered to youth to determine their level of involvement in video game playing indicate that key elements must be in place to ensure that a game will be played. These include action, strategy, and entertainment. Future investigation will examine the knowledge level of African American adolescent males of the nexus of certain concepts of climate change and health literacy. Climate change has significant implications for human health. This understanding will produce a scientifically based foundation for curricular and instructional decisions that include GBL. Results of this study will be used to design a video game concept and will contribute to the body of knowledge concerning environmental justice and empower individuals to make informed decisions about their own health and those they influence.

1.0 INTRODUCTION

Problem Based Learning (PBL) or Inquiry Based Learning (IBL) has long been recognized as effective practice in educational strategy [6],[8],[18],[22],[23],[24],[27]. Examples of PBL/IBL include a teacher's judicious use of simulation and role playing. Simulation is a recognized technique to allow students the chance to experience historical events and scientific procedures that might otherwise be unavailable to them. And role playing allows individuals to rehearse scenarios that are likely to occur in their actual employment or social environment. PBL/IBL seems to evolve naturally into Game Based Learning that incorporates the essential elements of simulation and role playing. However, if you design it, will they play it?

In this paper, the authors share the preliminary results of a study that examines the feasibility of GBL to increase the knowledge level of the relationship of climate change and health literacy among African American adolescent males. Specifically, Phase I of the study that addresses the level of involvement in video games among African American adolescent males will

be discussed. Additionally, possibilities for expansion of the concepts of modeling, simulation, and role playing in the design of computer games to increase health literacy among African American adolescent males will be suggested.

2.0 BACKGROUND

As noted, PBL/IBL role playing is not new to instruction in the K-12 arena. But incorporating technology into the process is a relatively new concept and one that is not universally practiced [20]. A critical question addressed in this study is "Can GBL increase the level of health literacy for individuals with a history of health disparities?" Teachers have begun to access video-modeling and its natural companions of simulation and role-playing to instruct students in academic, prevocational, and social skills. *Clover* is an example of a software program developed for middle school students in which they participate in simulations focusing on "social justice, honesty, and conflict resolution" [10].

Wehmeyer notes the success of using video-based modeling and simulation [28]. Examples of interactive videos in the health field include the simulations

of the Health Heroes series developed by Lieberman [11]. Scenarios address the topics of diabetes, asthma, surgery, bronchitis, and smoking. Single or multi-player formats are offered. Other games for health that have been developed include Internet-based virtual environments designed for behavior health care, wellness exercise gaming technology programs, and mobile applications to motivate and support healthy behavior change.

Today, climate change and its implications for healthy living is a focus of national concern. The National Institute of Health has noted that climate change poses a range of problems and has "potentially broad and significant implications for health" (NIH Notice, p.1) [16] in this century. Because of the pervasive concern of the effects of climate change on health [2], the topic is worthy of consideration for GBL. Additionally, health literacy and climate change are mission areas of the National Institute of Health (NIH), the Collaborative on Health and the Environment, the World Federation of Scientists, and the Climate and Health Literacy Consortium.

2.1 The Game's the Thing

Studies reveal the efficacy of Game Based Learning (GBL) on the acquisition, fluency, maintenance, and generalization of skill sets [3],[4],[19],[21]. Chuang et al (2009, p. 8) [1] note that "computer based gaming not only improves participants' fact level processes, but also promotes problem solving skills by recognizing multiple solutions for problems" [1]. This is consistent with current neuroscience that advocates the use of instructional strategies that engage both sides of the brain [7]. Although the efficacy of GBL is generally acknowledged as accepted educational practice, the motivation of students to use it cannot be assumed

[29]. The current White House initiative promoting healthy lifestyles that was targeted to African American urban youth relies on the print media of reading and writing. In her Blog posting, *Young Black Males, Learning, and Video Games*, Losh [12] comments on DiSalvo's work that emphasizes a different approach involving African American adolescent males in the design of video games. Additionally, Losh reports on Bruckman, whose work on gender and computer gaming, contends that looking at masculinity can provide an element of understanding in game design.

2.2 Check the Rock: Ensuring Players Want to Play

Before creating a GBL device to address climate change and its relevancy to health literacy among African American adolescent males, it is first necessary to determine the willingness of such individuals to access the game. What are the key components to promote usage? To this end, a GBL Usage Survey was devised. The survey was based on Whitton's work [29]. Subjects were students who were participating in a summer program that offers varied academic support services to assist students with transitioning from high school to college. While the survey was administered to males and females across all ethnicities in the program, only responses of those who self-reported as "male" and "African American" are presented. More than 100 students participated. Thirty three respondents reported as African American males. Not every participant responded to all twenty questions. Survey questions were divided into academic and affective domains. Table 1 depicts the frequency and percentages of a sampling of the survey responses.

Table 1 Video Gaming, Climate Change, Health Effects

Are you a video game player?	Yes	No
	N=28 (85%)	N=5 (15%)
How motivated would you be to learn, using video games?	Very much motivated	Not so motivated
	N=26 (81%)	N=6 (19%)
How many hours per day do you play video games during the school day?	One or less	Two or more
	N=15 (48%)	N=16 (52%)
How many hours per day on the weekend do you play video games?	One or less	Two or more
	N=8 (26%)	N=23 (74%)
What extent do you consider climate change to be an environmental problem?	Quite a problem	A medium problem or less
	N=25 (78%)	N=7 (22%)
What extent do you think climate change may affect human health?	It is very likely	It is a medium possibility or less
	N=27 (87%)	N=4 (13%)

Other information captured through the survey indicates that 84% of respondents play video games for entertainment, 10% play because friends or family play, and 2% play to learn more. Fourteen percent of those who noted they did not play games reported that games were not available to them. Eleven percent preferred “other leisure type activities”, and 7% thought playing video games “would waste your time.” Forty-seven percent chose an “action” game as a favorite type, 20%

chose “adventure and role playing”, 17% “simulation”, 13% “strategy”, and 3% “puzzle”. When asked to name games that were played the most often, respondents replied with examples of those in the sports or battle categories. Finally, participants were asked to describe what they have learned about the earth’s environment while playing a video game. Responses included concepts on pollution, nature and technology, weather patterns, destruction by humans, and “nothing.” Participants also noted that playing video games enabled them to learn a number of aspects about human health. These included damage to the body, energy, the importance of exercise, dehydration, and the role of medication. However, a majority of respondents replied with “nothing”, “none”, or “N/A”.

2.3 Capturing the Dream Team

Fantasy is an aspect of successful games. Fantasy employed in instruction can result in mastery of content as well as the thinking skill accessed for such mastery. The lives of students with environmental challenges as well as those who work with them are filled with reality checks, limitations, and boundaries. Gaming about one of those realities, maintaining good health, could afford such individuals not only an enhanced knowledge base about how climate change affects health, but also a degree of empowerment to those who may believe they often have “no say.”

In 1983, Williams [30] proposed that fantasy offered “. . . a new point of view and a new means of remembering information, and it can produce a ‘gut level’ understanding that goes much deeper and will be remembered longer than the verbal presentation of a text or lecture” (p. 117). Experiential learning that involves both hemispheres of the brain is vital to student success. Accessing potential experiences through

computerized role-playing games could afford individuals the opportunity to rehearse what has been presented as a viable model of a climate change/health literacy scenario. Players could likewise experience the results of choosing certain alternatives for overcoming encountered obstacles, as well as take action to help prevent the occurrence of events that cause health problems.

2.4 Putting It into Play

A variation of the concept of Massively Multi-player On-line Role-playing Game (MMORPG) could address the objective of making informed healthy choices. The persistent world of the game is the changing climate and its challenges to human health. And, different players (students, parents, professionals) would assume different roles.

Using Nintendo's "Hydlide" as a model, consider the following revision of the scenario [9]. "You play the role of (insert name) the (insert position: teacher, parent, student, health professional, etc.), a committed advocate for HEALTH armed with nothing but a diagnosis, prescription, and nobility of purpose. As the game begins, you sit in the center of a room surrounded by others who will from time to time attack your proposals. You will attack them back . . . and so the pursuit for HEALTH is on!" Progressing through the game, the player encounters common health related problems with alternative choices of solutions that could then yield either further "obstacles" or open passages for advancement toward a path to healthy living. These obstacles include (but are not limited to) a misdiagnosis, limited resources, no insurance, unhealthy environment, job demands, and personal obligations.

Additional instructional modules could include units on extreme weather events and changes in temperature and

precipitation. What are the effects on health of heat waves, hurricanes, tsunami, and severe drought? Flood destruction results in mold and mildew, contaminated food and water, and lack of shelter. Severe drought results in scarcities of food and water. What are the health outcomes of these occurrences? Heat stress, stroke, asthma, respiratory disease, malnutrition, starvation, and other concerns resulting from air pollution, disease carriers and contaminants are examples of how climate change can harm the public's health. A primary learning objective is that game players would acquire a knowledge of how a lack of clean air and water, as well as unhealthy living conditions are made worse by adverse effects of climate change. Another learning objective is that a player would gain a skill set to make informed decisions about health and its relationship to the environment.

Analysis of the findings of this first phase of the study should provide suggestions for how the game should be crafted to target African American adolescent males successfully. It is anticipated that these young men will also be involved in the design of the game. Results of a climate change and health literacy survey administered to the same group of participants focused on gathering information on what youth know about the relationship between climate change and their health and about some of the personal choices they make. These findings will also be incorporated into the game design. Investigators are collaborating with personnel in the University's Center for Gaming and Simulation about the development of the actual game. This collaboration will determine the structure of future interviews and surveys to be administered to groups and analyzed to identify variables. Such collaboration will also provide technical support and

guidance for the development of the game.

What approach should be used with adolescent African American males to ensure participation in health literacy GBL? Whitton [29] has concluded that students will access games to learn if they are convinced that using such games is the most effective way to learn the material. Her study found that there was no evidence to surmise that playing games was motivating in itself to learn. The players needed to believe that such "play" was the best way to learn the content. The investigators' approach is the integration of the health and environmental science standards of the curriculum required by the State Department of Education into a user friendly serious gaming format.

To engage learners in GBL, a number of factors must be considered. A primary consideration is the recognition of the human factor. Teachers who use GBL must be culturally aware and culturally responsive to the need for enthusiasm about the GBL process. They must be available for face to face affirmation when requested, and they must acknowledge that their African American adolescent male students often hold fast to the idea that "the body is first and foremost in identity" [12]. Investigators will continue to explore what needs to be different about the game to target African American males. This might entail the provision for a passion for sports; visual, tactile, and kinesthetic experiences; emphasis on right brained approaches; a mechanism for family engagement; and an emphasis in body centric athleticism.

3.0 DISCUSSION

The PEW Internet and American Life Project (2009) reports that 99% of boys and 94% of girls ages 12-17 access technology as a social activity. Included

in the PEW Project is Lenhart's (2008) [10] study, "Teens, Video Games, and Civics". She further notes that of the various game genres played by teens, 49% play simulations, 59% play strategy games, and 36% engage in role playing. Of particular note is the finding that 78% of the respondents reported that they "frequently" or "sometimes" witnessed kindness and helpfulness to other gamers by those who were playing. The study also found that general frequency of use was not related to civic and social isolation but actually could lead to greater levels of civic engagement [10].

The provision of an interactive learning tool to increase health literacy among African American adolescent males could allow them to progress in their self-advocacy and social skills while participating in an age appropriate activity. Teachers of a health literacy curriculum might determine that access to gaming devices will become part of a student's program of study. Additionally, Fox [5] contends that becoming a serious gamer could be a transition goal for students and could serve as a social bridge for many students.

The concept of a collaborative virtual environment (CVE) to learn and practice skills has proven successful with various student populations [14]. Simply put, a CVE creates a virtual space in which humanoid avatars can meet and interact. A viable CVE contains some key features. These include "shared context, awareness of others, negotiation, communication, and flexible and multiple viewpoints" [14]. The creation of a health literacy game using CVE could enhance cooperative learning and decision making.

Educators have come to realize the power of social context in affecting the development of ideas [26]. In his discussion of the future of modeling and simulation, Oren (2002) [17] notes that

"In zero-sum games, competition is essential. In non-zero sum games, cooperation is essential. Simulation can be useful in teaching the value and practice of cooperation" (p.121). This is also a valuable asset for social groups. The ability to collaborate with health professionals about one's health status and care is necessary for everyone. It is vital for those numbered among the vulnerable populations. An MMO social group can provide real world simulations and advance the agenda for individuals with a history of being disenfranchised.

4.0 CONCLUSION

The reauthorization of the Elementary and Secondary Education Act (NCLB, 2002) [25] has prodded school districts to develop instructional approaches to ensure that no child is left behind. Content must be delivered in a way that is cognitively and affectively comparable to the diverse backgrounds and learning styles of today's learners. One such approach is the use of Game Based Learning (GBL). A critical question emerging from results of the GBL Usage Survey is how to construct an instructional game that will be used by African American adolescent males. Results suggest that a number of major elements must be considered to craft an efficient and effective game to enhance learning. These include components of action and strategy. Additionally, the design must be entertaining.

Authors of this paper have sought to examine the concept of employing computer assisted technology to enhance the knowledge acquisition and skill application of individuals involved in learning how climate change may impact health. It is suggested that such an approach is feasible and fulfills not only the letter of the law but the spirit of NCLB. However, as worthy as this endeavor might be, critical to the process of research and development in

GBL for attaining full access to any knowledge construct such as health literacy is the need for an approach that blends competence with compassion. "It will not suffice simply to design a knob to match human requirements; the whole system must be designed to take into account the whole human being" [15].

The use of GBL as a method to provide effective instruction is a valuable objective for teaching the relationship between climate change and health. It can ensure that individuals will have a voice in the planet's growth. A voice that ensures the world is left "a little more peaceful, healthier, and richer in biotic life than when we entered it" (Maathai, 2010, p. 194) [13]. Enhancing the experience of African American adolescent males in game based learning could contribute to overcoming the critical global challenge of maintaining good health in a changing climate.

REFERENCES

1. Chuang, T. & Chen, W. (2009) "Effect of Computer Based Video Games on Children: An Experimental Study", *Educational Technology and Society*, vol.12, no.2, pp. 1-10.
2. Costello, A., Abbas, M., Allen, Adriana, et al. (2009) "Managing the Health Effects of Climate Change", *The Lancet*, vol.373, pp.1693-1733.
3. Cohen, R.B. & Bradley, R.H. (1993) "The Use of Simulation Games to Create Real Worlds for Special Education Students", *Reading Improvement*, vol. 15, no. 4, pp. 276-278.
4. DiPietro, M., Ferdig, R.E. Boyer, J. & Black, E. (2007) "Towards a Framework of Understanding Electronic Educational Gaming" *Journal of Educational Multimedia and Hypermedia*, vol. 16, no. 3, pp. 225-228.
5. Fox, C. (2007) "Serious Games Have Serious Possibilities" *Special Education Law Blog*

- http://specialledlaw.blogs.com/home/2007/02/serious_games.html. Retrieved 6/30/09.
6. Heward, W.L. (2009) *Exceptional Children: An Introduction to Special Education*. Pearson: Upper Saddle River, NJ.
 7. Jensen, E. (2000) *Brain-based Learning*. The Brain Store: San Diego, CA.
 8. Koch, J. (2009) *So You Want to Be a Teacher?: Teaching and Learning in the 21st Century*. Houghton Mifflin Company: New York.
 9. Lampton, C. (1991) *Nintendo: role-Playing Games*. The Millbrook Press: Brookfield, CT.
 10. Lenhart, A. (2008) "Teens, Video Games, and Civics", *Games, Learning and Libraries Conference*, November 2, 2008.
 11. Lieberman, D.A. (1997) *Interactive Video Games for health Promotion: Effects on Knowledge, Self-efficacy, Social Support, and Health. Health Promotion and Interactive Technology: Theoretical Applications and Future Directions*. Lawrence Erlbaum Associates, Inc; Publishers: Mahwah, NJ.
 12. Losh, L. (2011, February 17) Re: "Young Black Males, Learning, and Video Games", *DML Web log message*. Retrieved from <http://dmlcentral.net/node/3660>.
 13. Maathai, W. (2010) *Replenishing the Earth: Spiritual Values for Healing Ourselves and the World*. Doubleday: New York.
 14. Moore, D., Cheng, Y., McGrath, P., & Powell, N.J. (2005) "Collaborative Virtual Environment Technology" and People with Autism", *Focus on Autism and Other developmental Disabilities*, http://www.accessmylibrary.com/coms2/sum_mary-0286-123-02045_ITM. Retrieved 7/22/09.
 15. Norman, D. A. (1982) *Learning and Memory*. W. H. Freeman Press: New York.
 16. Notice to Highlight Current Funding Opportunities that Promote Research on the Human Health Effects of Climate Change. NOT-TW-10-008. June 24, 2010, pp.1-5.
 17. Oren, T.I. (2002) "Future of Modeling and Simulation: Some Development Areas", *Proceedings of the 2002 Summer Computer Simulation Conference*.
 18. Petersen, M.J. & Hittie, M.M. (2010) *Inclusive Teaching: The Journey Towards Effective Schools for All Learners*. Pearson: Upper Saddle River, NJ.
 19. Reese, D.D. (2007) "First Steps and Beyond: Serious Games as Preparation for Future Learning", *Journal of Educational Multimedia and Hypermedia*, vol. 16, no. 3, pp. 283-300.
 20. Rice, J. (2007) "New Media Resistance: Barriers to Implementation of Computer Video Games in the Classroom" *Journal of Educational Multimedia and Hypermedia*, vol. 16, no. 3, pp. 249-261.
 21. Rice, J. (2007) "Assessing Higher Order Thinking in Video Games", *Journal of Technology and Teacher Education*, vol. 15, no. 1, pp.87-100.
 22. Sadker, D.M. & Zittleman, K.R. (2010) *Teachers, Schools, and Society*. McGraw Hill: New York.
 23. Salend, S.J. (2008) *Creating Inclusive Classrooms: Effective and Reflective Practices*. Pearson: Upper Saddle River, NJ.
 24. Silver, H.F., Strong, R.W., & Perini, M.J. (2000) *So Each May Learn: Integrating Learning Styles and Multiple Intelligences*. ASCD: Arlington, VA.
 25. United States Department of Education (2001) PL 107-110. No Child Left Behind Act.
 26. Vgotsky, L. (1962) *Thought and Language*. MIT Press: Cambridge, MA.
 27. Villa, R.A., Thousand, J.S., Stainback, W. & Stainback, S. (1992) *Restructuring for Caring and Effective Education: An Administrative Guide to Creating Heterogeneous Schools*. Paul H. Brooks Publishing, Baltimore, MD.
 28. Wehmeyer, M. (2007) *Promoting Self-Determination in Students with Developmental Disabilities*. Guilford Press: New York.
 29. Whitton, N. (2007) "Motivation and Game Based Learning", *ascilite Proceedings Singapore*, pp. 1063-1067.
 30. Williams, L.V. (1983) *Teaching for the Two-Sided Mind*. Simon and Schuster, Inc.: New York.

5.16 What Research Says About Competition and Learning

What Research Says About Competition and Learning

Richard Wainess
UCLA / NCRESST
wainess@ucla.edu

Abstract: Proponents of the use of video games for learning often cite motivation as a reason for using games. A number of motivational characteristics have been attributed to games, including the motivation of competition. While research does support the contention that motivation is good for learning, it doesn't necessarily support competition as good for learning. There are several key factors that determine whether competition is helpful or harmful. One is whether the student is winning or losing. There is strong research evidence that losing and negative feedback reduce the likelihood of learning by a) increasing the desire to withdraw from the task, which reduces the amount of mental effort a learner will exert, and b) placing negative, task irrelevant thoughts in the learner's working memory which reduces working memory capacity and effectiveness. Competitive learning environments can also change the learner's goal from pursuit of success to avoidance of failure, where the learner is less concerned with learning and more concerned with the prospect of failing. A third factor is the learner's goal orientation. There are four types of goal orientation. Each influences how a learner reacts to public display of ability (or inability) and whether the learner is more focused on gaining knowledge and skills or displaying knowledge and skills. The research evidence suggests that, while competition may be good for some learners or some learning situations, it is generally not a fruitful strategy for learning, including learning from games.

1.0 INTRODUCTION

The mantra of advocates for video games in education seems to be "games are good for learning because they're motivating" (Dickey, 2005; Gee, 2003, 2006; Simpson, 2005). This mantra includes the belief that motivation leads to learning; or to some researchers, *guarantees* learning. Much of this connection between games, motivation, and learning dates back to the work of Malone (1981) and colleagues (e.g., Malone & Lepper, 1987) where a number of constructs were identified as providing the motivational appeal of games and the implication that these constructs and their related motivational characteristics were beneficial to learning and, ultimately, learning from games. However, empirical evidence hasn't necessarily supported those conclusions. For example,

competition has been cited as a motivational characteristic of games (Malone, 1981), yet research on competition and learning from games suggests that competition can be detrimental to learning (e.g., Coutinho & Neuman, 2008; Linnenbrink, Ryan, & Pintrich, 1999; Tauer & Harackiewicz, 1999).

2.0 FRAMEWORK OF MOTIVATION AND GAME-BASED LEARNING

Wainess (2008) presented a framework to guide research on motivation and game-based learning. The framework was based on an extensive body of empirical research linking motivation (the independent variable; IV) to learning (the dependent variable; DV) through various mediating and moderating variables.

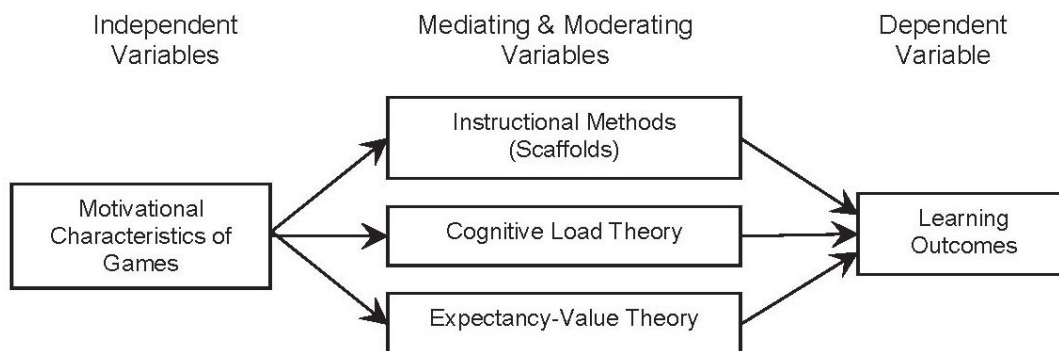


Figure 1 Simplified Model of Motivation and Learning IVs, MVs, and DV.

Fig. 1 depicts a simplified version of the model that supports the Wainess (2008) research framework. The relationship between various motivational characteristics of games (the IVs) and learning outcomes (the DV) are moderated and mediated by three variable groups: *instructional methods*, *cognitive load theory*, and *expectancy-value theory*. The IVs comprise the motivational constructs originally proposed by Malone (1981) along with motivational constructs added by later research (e.g., Dickey, 2005; Gee, 2003, 2006; Simpson, 2005). The MVs are those most relevant to the IVs and to learning from game-based environments. They are the constructs that most often appear in research related to the IVs and the DV. The

DV—learning outcomes—is an ordinal variable with a range of potential values from low (minimal) learning to high (substantial) learning.

The variables and, ultimately, the model are based on empirical data indicating correlations and causal connections among the various IVs and MVs and their impact on the DV. A mediating variable explains how the relationship between an independent and dependent variable occurs (retrieved July 27, 2007 from http://en.wikipedia.org/wiki/Mediator_variable). Moderating variables describe the conditions under which an independent variable exerts its effect on a dependent variable.

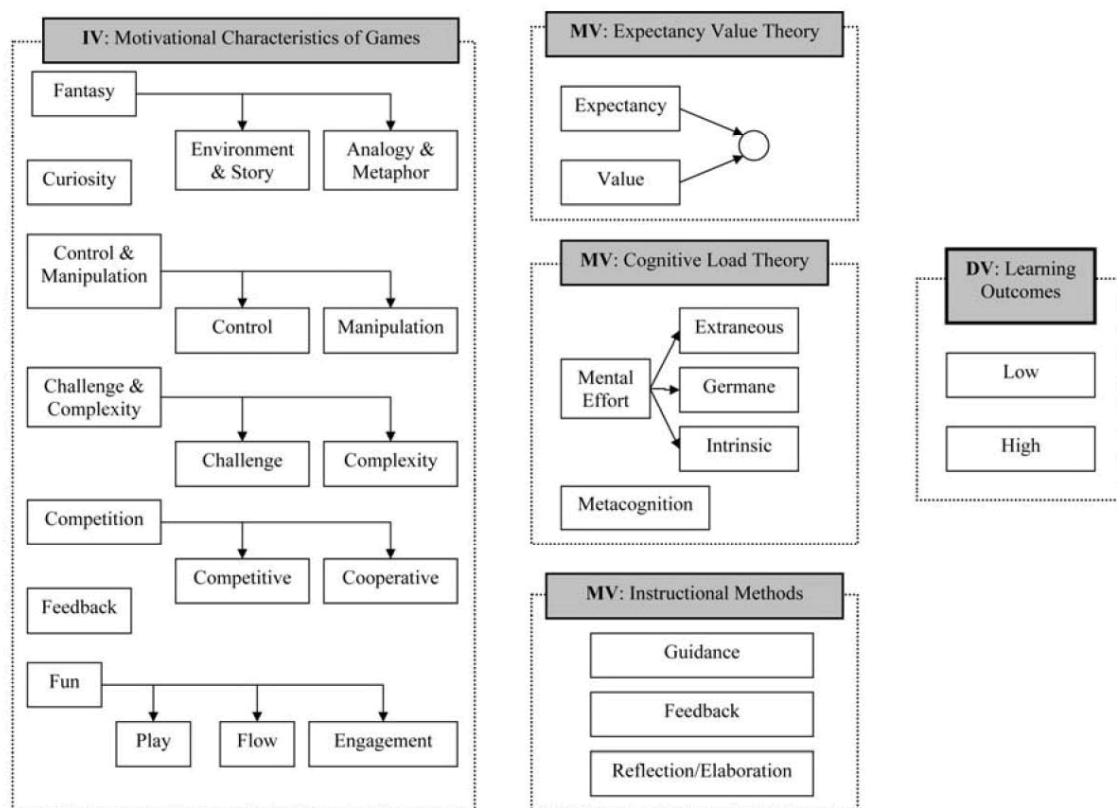


Figure 2 Model of the Relationship Between Motivational Constructs and Game-Based Learning Outcomes.

Fig. 2 is an expanded model of the overview depicted in Fig. 1. In the expanded model, seven motivational characteristics (or

constructs) are shown on the left. Five of the seven are further analyzed into subconstructs. In the center portion of the

model, each of the three mediating or moderating variables has been analyzed into their component constructs. The two constructs under expectancy value theory (*expectancy* and *value*) are the two constructs that comprise that theory (Eccles & Wigfield, 2002). For the other two moderating/mediating variables (*cognitive load theory* and *instructional methods*), the constructs shown represent only some of those associated with the variable. In both instances, the constructs listed are those most relevant to research findings related to the IV constructs of motivation and the DV of learning outcomes (learning from games). For a complete explanation of the model, see Wainess (2008).

3.0 RELATIONSHIP OF COMPETITION TO LEARNING

While the assumption is that motivation always leads to learning, research suggests that this assumption may not be true. A number of researchers offer evidence that motivation does not guarantee learning (e.g., Brougere, 1999; Druckman, 1995; Salas, Bowers, & Rhodenizer, 1998). Salomon (1983) even contended that a positive attitude can actually indicate less learning. Various motivational constructs have the ability to either positively or negatively affect expectancy-value, and working memory, both of which have been shown to impact learning (Bong, 2001; Elliman, Green, Rogers, & Finch, 1997).

Competition is an example of potentially negative learning outcomes from a motivational construct. A substantial body of research indicates that *goal orientation* (Linnenbrink et al., 1999, Tauer & Harackiewicz, 1999), *reward structure preference* (Brown, Cron, & Slocum Jr., 1998; Porter, Bird, & Wunder, 1990-1991; Slavin, 1977), an win/lose state (Ames, Ames, & Felker, 1977) are determinants of whether competition is motivating or demotivating. In terms of *expectancy-value theory*, most research points primarily to the *value* part of expectancy-value theory as a

dominant moderator for competition and learning outcomes (Coutinho & Neuman, 2008; Linnenbrink et al., 1999; Tauer & Harackiewicz, 1999), while a lesser amount points to the *expectancy* part of expectancy-value theory (Ames et al., 1977; Brown et al., 1998). In terms of *cognitive load theory* the research points to working memory issues (Coutinho & Neuman, 2008; Linnenbrink et al. 1999) as a moderator for learning outcomes. Other research links goal orientation as a moderator for learning outcomes (Covington, 1984, Tauer & Harackiewicz, 1999). Next we examine how each of these three moderating groups can negatively affect learning. They will be examined in the following order.

- Expectancy-Value theory
- Cognitive Load Theory
- Goal Orientation

4.0 EXPECTANCY VALUE THEORY

Expectancy-value theory proposes that the probability of behavior depends on the value of the goal and the expectancy of obtaining that goal (Coffin & MacIntyre, 1999). If a learner values an experience and has some expectation of success, then he or she might be willing to engage in that experience. If, however, the learner does not value the experience or does not believe he or she can succeed, the learner will disengage from the experience and reduce or halt mental effort toward that experience (Clark, 1999; Eccles and Wigfield, 2002).

5.0 COGNITIVE LOAD THEORY

Cognitive load theory (Baddeley, 1986; Brunken, Plass, & Leutner, 2003; Kirschner, Sweller, & Clark, 2006; Mayer & Moreno, 2003; Mousavi, Low, & Sweller, 1995; Paas, Tuovinen, & Tabbers, 2003; Renkl & Atkinson, 2003) is concerned with the development of instructional methods aligned with learners' limited cognitive processing capacity. Researchers have proposed that working memory limitations can have an adverse effect on learning (Sweller & Chandler, 1994; Yeung, 1999). Miller (1956) described a working memory

capacity of between five and nine *chunks* of information. Cognitive load researchers have identified up to three types of cognitive load that can be imposed on working memory; *intrinsic cognitive load*, *germane cognitive load*, and *extraneous cognitive load* (see Fig. 2). Of particular importance to this paper is extraneous cognitive load (Harp & Mayer, 1998; Renkl & Atkinson, 2003; Schraw, 1998). Unnecessary items in working memory are globally referred to as extraneous cognitive load. Any thought, including extraneous thought, reduces available working memory (Elliman et al., 1997).

6.0 GOAL ORIENTATION

There are two major goal orientations: *mastery* goal orientation and *performance* goal orientation. Those with a *mastery goal orientation* focus on the development of new skills or competencies and on task mastery (Elliott & Dweck, 1988; VandeWalle, Brown, Cron, & Slocum Jr., 1999). They view a challenging task as an opportunity for growth and development. They persist, escalate effort, engage in solution-oriented self instruction, and enjoy the challenge (VandeWalle et al., 1999).

Those with a *performance goal orientation* focus on demonstrating competence relative

to others and seek favorable judgments (Linnenbrink et al., 1999; VandeWalle et al., 1999). They view a challenging task as a threat because of the risk of failure. In challenging situations, they withdraw from the task, make negative ability attributions, and report decreased interest in the task.

The two major goal orientation categories, *mastery* and *performance*, can be further analyzed into *approach* and *avoidance*: That is, *mastery-approach*, *mastery-avoidance*, *performance-approach*, and *performance-avoidance*. According to Coutinho and Neuman (2008), those with a *mastery-approach* goal orientation work to attain positive results (e.g., acquiring knowledge or skills), while those with a *mastery-avoidance* goal orientation work to avoid losing skills or becoming incompetent. According to Brophy (2005), those with a *performance-approach* goal orientation want to demonstrate their ability relative to others by outperforming them and publicly displaying their knowledge or skills, while those with a *performance-avoidance* goal orientation want to avoid public display of inability or incompetence. People with performance goals orientations tend to avoid challenges and obstacles where the risk of failure is high (Coutinho & Neuman, 2008)

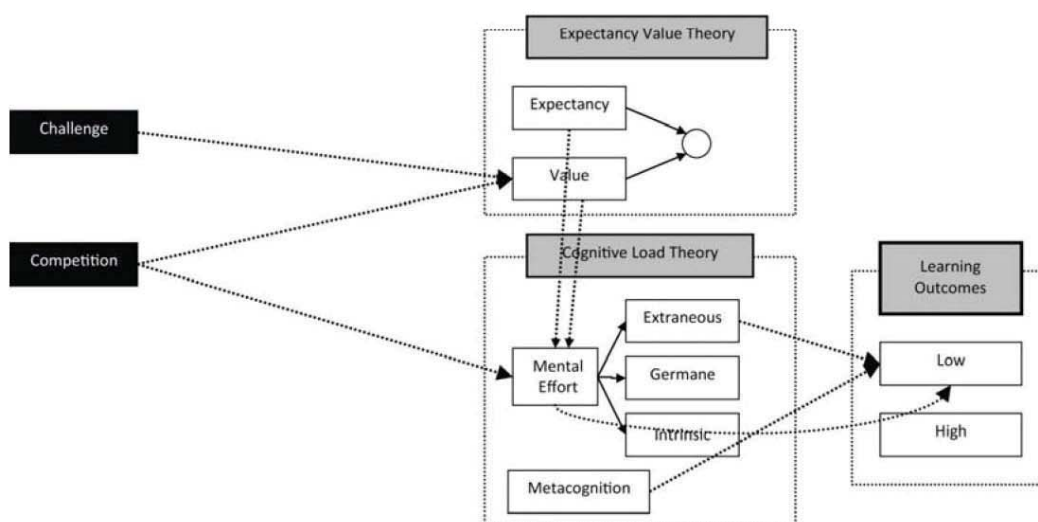


Figure 3 Relationship between Competition and Challenge, Key Moderating and Mediating Variables, and Learning Outcomes

Fig. 3 shows the paths for how competition can negatively affect learning outcomes. Fig. 3 shows the relevant sections of the model for motivation and game-based learning (see Fig.2). In Fig. 3, the left components are competition and challenge. *Competition* refers to two or more parties striving for a goal that cannot be shared (retrieved June 11, 2010 from <http://en.wikipedia.org/wiki/Competition>). *Challenge*, was defined by Berube,

Severynse, Jost, Ellis et al. (2001) as a “requirement for full use of one’s abilities or resources” (p. 185). This definition of challenge suggests that challenge is a necessary ingredient of competition. Fig. 3 indicates that both competition and challenge can negatively affect the value a learner places on the learning task, which will in turn affect the amount of mental effort the learner places on the task and, ultimately, the amount of learning that occurs.

7.0 HOW COMPETITION AFFECTS LEARNING

There are two ways competition (and challenge) can harm learning from games. They can 1) negatively affect a learner’s perceived value of the learning task and 2) reduce a learner’s working memory. Each of these two potential interactions is discussed in order.

- Reduction in perceived value
- Reduction in working memory

7.1 Reduction in Perceived Value

Fig. 4 adds a number of moderating and mediating variables between competition and challenge and value. It is these additional variables that explain the reduction in value that can occur from competition and challenge. Fig. 4 shows that competition, combined with losing and the negative feedback associated with losing can affect competence valuation and enjoyment.

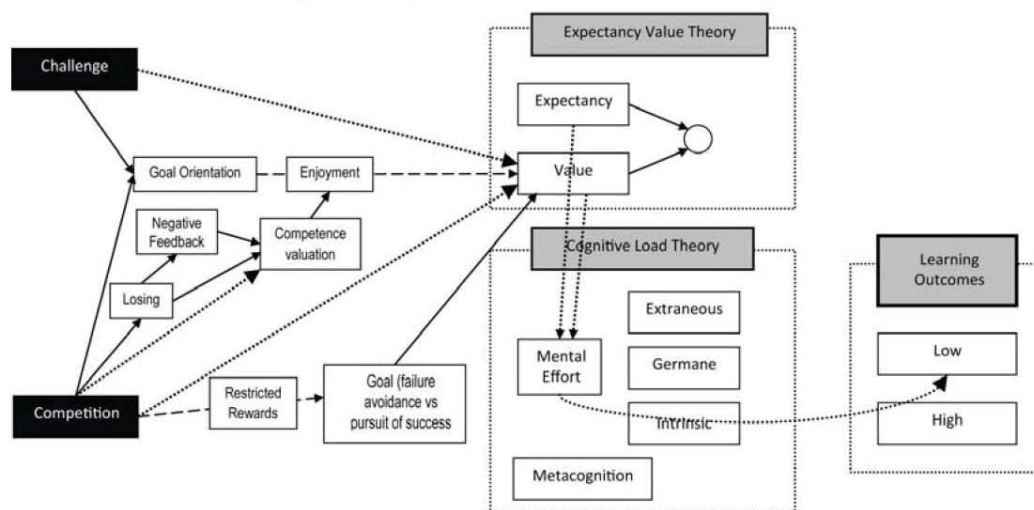


Figure 4 Competition and Challenge, Expectancy Value Theory, and Learning Outcomes

Competence valuation refers to how competent a person feels about his or her abilities or performance. According to Tauer and Harackiewicz (1999), competence valuation is invoked at the outset of competition and is a significant predictor of enjoyment. Competition can amplify competence valuation and, in the case of

losing and negative feedback, amplify negative competence valuations. In a study by Ames et al. (1977), fifth grade boys in a competitive environment exhibited negative self-attribution affects associated with losing, while those in the non-competitive environment exhibited no such effect. When competence valuation is reduced,

enjoyment is also affected, which reduces value reduces mental effort, and reduces learning. *Meaningful learning* is defined as deep understanding of the material (Moreno & Mayer, 2003). Meaningful learning requires *mental effort* (Davis & Wiedenbeck, 2001; Mayer, 1981) which Salomon (1983) described as the depth or thoughtfulness a learner invests in processing material. When mental effort is reduced, meaningful learning is negatively affected. Another potential negative effect of competition is it can change a learner's goals to avoidance of failure, rather than pursuit of success (Covington, 1984). This in turn can reduce the value of the learning, reduce mental effort, and reduce learning (see Fig. 4).

7.2 Reduction in Working Memory

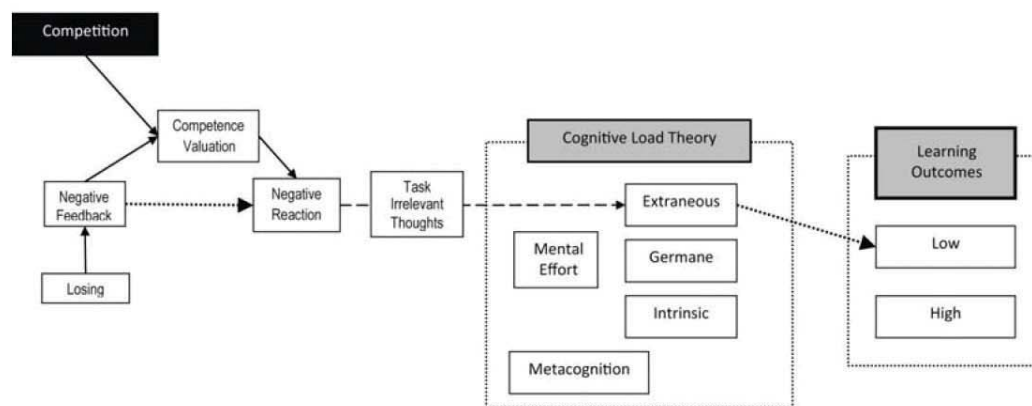


Figure 5 Relationship between Competition, Working Memory, and Learning Outcomes

9.0 DISCUSSION AND CONCLUSIONS

While competition has been cited as a motivational benefit of games, research indicates that competition can actually demotivate players when the game is used in a learning context. Competition heightens competence valuation, but heightened competence valuation inflates the negative effects of losing and of negative feedback. Competition also changes learner's goals from pursuit of success to avoidance of failure. That change from pursuit to avoidance reduces the value learners place on the learning task, which in turn reduces the amount of

Working memory is limited to only 5 to 9 thoughts at a time (Miller, 1956) and any thought not relevant to the current learning task unnecessarily burdens the learner's already limited working memory capacity (Ellis & Ashbrook, 1988; Linnenbrink et al., 1999). The irrelevant thoughts alter the allocation of resources and interfere with the central executive process and the articulatory loop (Elliman et al., 1997). Fig. 5 shows that competition combined with the negative feedback and the negative reactions that come from losing can introduce task irrelevant thoughts about ones performance. Those thoughts become extraneous cognitive load which unnecessarily reduces available working memory and, ultimately, reduces learning.

mental effort they apply, resulting in less learning. Losing and negative feedback also introduce negative thoughts; thoughts that are irrelevant to the actual task of learning. Because working memory capacity is small, any unnecessary burden on working memory has a negative effect on learning; task irrelevant thoughts reduce working memory capacity.

Fig. 2 showed a model of the relationship between motivational constructs and game-based learning outcomes. Fig. 3 showed the portion of the model related to how

competition (and challenge) affect learning. They do so by affecting value and by affecting working memory (Figs. 4 and 5). This paper explains why value is affected and why working memory is affected. The mediating variables in Table 1 (value and

working memory) are from the model represented in Fig. 2. The independent and moderating variables in Table 2 are those involved in the explanation of how value and working memory are affected.

Table 1: The IVs, MVs, and DVs of Competition and Learning

Independent Variables	Moderating Variables	Mediating Variables	Dependent Variable
• Goal orientation • Reward Structure	• Structure Preference • Competence Valuation	• Value • Working Memory	• Learning Outcomes • Video Game Performance

It must be noted that, while the finding of this paper strongly support the notion that competition should not be used in game-based-learning, there are some instances where competition may not be harmful and might even be beneficial. For example, some learning situations may involve learners who are not so averse to competition. For some learners, competition and its related constructs (e.g., competence valuation, enjoyment) can be positive, or at least, not negative. Tauer and Harackiewicz (1999) found that the effect of competition is moderated by goal orientation, which matches the findings of a study by Linnenbrink et al. (1999). Tauer and Harackiewicz's found that while learners low in achievement motivation (those with a performance-avoidance goal orientation) reacted negatively to competition, those high in achievement motivation (those with a performance-approach goal orientation) did not. Brown et al. (1998) also found that high trait competitiveness combined with a competitive environment lead to high goal setting.

Additionally, Coutinho and Neuman (2008) found that learners can have multiple goal orientations operating independently of each other, such as a mastery orientation with regards to acquiring new knowledge

but a performance orientation with regards to test performance. Slavin (1977) suggested that a blend of cooperative and competitive or cooperative and individual reward structures might be a valuable approach for academic achievement.

Without knowing learners' goal orientations, it would be difficult to determine when or if competition is a viable learning strategy. Therefore, we recommend not including competition unless it is a necessary component of learning. For example, competition is an integral component of many sports and of marketing and sales. It might be difficult to teach all elements of these two domains unless competition were included at some point in the learning process.

10.0 ACKNOWLEDGMENTS

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305C080015, as well as grant number N00014-08-C-0563 from the Office of Naval Research with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The findings and opinions expressed here do not necessarily reflect the positions or policies of the National Center for Research on Evaluation, Standards, and Student Testing, the Institute of Education Sciences, the U.S. Department of Education, or the Office of Naval Research.

11.0 REFERENCES

- [1]. Ames, C., Ames, R., & Felker, D. W. (1977). Effects of Competitive Reward Structure and Valence of Outcome on Children's Achievement Attributions. *Journal of Educational Psychology*, 69(1), 1-8.
- [2]. Baddeley, A. D. (1986). *Working Memory*. Oxford, England: Oxford University Press.
- [3]. Berube, M. S., Severynse, M., Jost, D. A., Ellis, K., Pickett, J. P., Previte, R. E., et al. (Eds.) (2001). *Webster's II: New College Dictionary*. Boston, MA: Houghton Mifflin Company.
- [4]. Bong, M. (2001). Between- and Within-Domain Relationships of Academic Motivation Among Middle and High School Students: Self-efficacy, Task-value, and Achievement Goals. *Journal of Educational Psychology*, 93(1), 23-34
- [5]. Brophy, J. (2005). Goal Theorists Should Move on from Performance Goals. *Educational Psychologist*, 40(3), 167-176.
- [6]. Brougere, G. (1999, June). Some Elements Relating to Children's Play and Adult Simulation/Gaming. *Simulation & Gaming*, 30(2), 134-146.
- [7]. Brown, S. P., Cron, W. L., & Slocum Jr., J. W. (1998). Effects of Trait Competitiveness and Perceived Intraorganizational Competition on Salesperson Goal Setting and Performance. *The Journal of Marketing*, 62(4), 88-98.
- [8]. Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist* 38(1), 53-61.
- [9]. Clark, R. E. (1999). The CANE Model of Motivation to Learn and to Work: A Two-stage Process of Goal Commitment and Effort [Electronic Version]. In J. Lowyck (Ed.), *Trends in Corporate Training*. Leuven, Belgium: University of Leuven Press.
- [10]. Coffin, R. J., & MacIntyre, P. D. (1999). Motivational Influences on Computer-related Affective States. *Computers in Human Behavior*, 15, 549-569.
- [11]. Coutinho, S. A., & Neuman, G. (2008). A Model of Metacognition, Achievement Goal Orientation, Learning Style and Self-efficacy. *Learning Environments Research*, 11, 131-151.
- [12]. Davis, S., & Wiedenbeck, S. (2001). The Mediating Effects of Intrinsic Motivation, Ease of Use and Usefulness Perceptions on Performance in First-time and Subsequent Computer Users. *Interacting with Computers*, 13, 549-580.
- [13]. Dickey, M. D. (2005). Engaging by Design: How Engagement Strategies in Popular Computer and Video Games Can Inform Instructional Design. *Educational Technology Research & Development*, 53(2), 67-83.
- [14]. Druckman, D. (1995). The Educational Effectiveness of Interactive Games. In D. Crookall & K. Aria (Eds.), *Simulation and gaming across disciplines and cultures: ISAGA at a watershed* (pp. 178-187). Thousand Oaks, CA: Sage
- [15]. Eccles, J. S., & Wigfeld, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, 53, 109-132.
- [16]. Elliman, N. A., Green, M. W., Rogers, P. J., & Bretel, M. Finch (1997). Processing-efficiency Theory and the Working-memory System: Impairments Associated with Sub Clinical Anxiety. *Personality and Individual Differences* 23(1), 31-35.
- [17]. Elliott, E. S., & Dweck, C. S. (1988). Goals: An Approach to Motivation and

- Achievement. *Journal of Personality and Social Psychology*, 54(1), 5-12.
- [18]. Ellis, H. C., & Ashbrook, T. W. (1988). Resource Allocation Model of the Effects of Depressed Mood State on Memory. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior* (pp. 25-43). Toronto, Canada: Hogrefe.
- [19]. Gee, J. P. (2003) *What Video Games Have to Teach us About Learning and Literacy*. New York: Palgrave Macmillan.
- [20]. Gee, J. P. (2006, August). *Are Video Games Good for Learning*. Keynote address presented at the Curriculum Corporation 13th Annual Conference, Adelaide, Australia.
- [21]. Harp, S. F., & Mayer, R. E. (1998). How Seductive Details do their Damage: A Theory of Cognitive Interest in Science Learning. *Journal of Educational Psychology*, 90(3), 414-434.
- [22]. Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-based, Experiential, and Inquiry-based Teaching. *Educational Psychologist*, 41(2), 75-86.
- [23]. Linnenbrink, E. A., Ryan, A. M., Pintrich, P. R. (1999). The Role of Goals and Affect in Working Memory Functioning. *Learning and Individual Differences* 11(2), 213-230.
- [24]. Malone, T. W. (1981). What Makes Computer Games Fun? *Byte*, 6(12), 258-277.
- [25]. Malone, T. W., & Lepper, M. R. (1987). Making Learning Fun: A Taxonomy of Intrinsic Motivation for Learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3. Cognitive and affective process analyses* (pp. 223-253). Hillsdale, NJ: Lawrence Erlbaum.
- [26]. Mayer, R. E. (1981). A psychology of How Novices Learn Computer Programming. *Computing Surveys*, 13, 121-141.
- [27]. Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43-52.
- [28]. Miller, G. A. (1956). The Magical Number, Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.
- [29]. Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology*, 87(2), 319-334.
- [30]. Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), 63-71.
- [31]. Porter, D. B., Bird, M. E., & Wunder, A. (1990-1991). Competition, Cooperation, Satisfaction, and the Performance of Complex Tasks Among Air Force Cadets. *Current Psychology: Research & Reviews*, 9(4), 347-354.
- [32]. Renkl, A., & Atkinson, R. K. (2003). Structuring the Transition from Example Study to Problem Solving in Cognitive Skill Acquisition: A Cognitive Load Perspective. *Educational Psychologist*, 38(1), 13-22.
- [33]. Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training. *The International Journal of Aviation Psychology*, 8(3), 197-208.

- [34]. Salomon, G. (1983). The Differential Investment of Mental Effort in Learning from Different Sources. *Educational Psychology*, 18(1), 42-50.
- [35]. Schraw, G. (1998). Processing and Recall Differences Among Seductive Details. *Journal of Educational Psychology*, 90(1), 3-12.
- [36]. Simpson, E. S. (2005). What Teachers Need to Know About the Video Game Generation. *TechTrends*, 49(5), 17-22.
- [37]. Slavin, R. E. (1977). Classroom Reward Structure: An Analytical and Practical Review. *Review of Educational Research*, 47(4), 633-650.
- [38]. Sweller, J., & Chandler, P. (1994). Why Some Material is Difficult to Learn. *Cognition and Instruction*, 12, 185-233.
- [39]. Tauer, J. M., & Harackiewicz, J. M. (1999). Winning Isn't Everything: Competition, Achievement Orientation, and Intrinsic Motivation. *Journal of Experimental Social Psychology*, 38, 209-238.
- [40]. VandeWalle, D., Brown, S. P., Cron, W. L., & Slocum Jr., J. W. (1999). The Influence of Goal Orientation and Self-Regulation on Sales Performance: A Longitudinal Field Test. *Journal of Applied Psychology*, 84(2), 249-259.
- [41]. Wainess, R. (2008). A Framework for Research on Motivation and Game-based Learning. *Proceedings from the Future and Reality of Games conference, Vienna, Austria, held October 17-19, 2008*.
- [42]. Yeung, A. S. (1999). Cognitive Load and Learner Expertise: Split Attention and Redundancy Effects in Reading Comprehension Tasks with Vocabulary Definitions. *Journal of Educational Media*, 24(2), 87-102.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 01-03-2012		2. REPORT TYPE Conference Publication		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Selected Papers Presented at MODSIM World 2011 Conference & Expo			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Pinelli, Thomas E.; Bullock, Leanna S. (Compilers)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER 736466.01.04.07.08		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199			8. PERFORMING ORGANIZATION REPORT NUMBER L-20111		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001			10. SPONSOR/MONITOR'S ACRONYM(S) NASA		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) NASA/CP-2012-217326		
12. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 66 System Analysis and Operations Research Availability: NASA CASI (443) 757-5802					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Selected papers from MODSIM World 2011 Conference & Expo are contained in this NASA Conference Publication (CP). MODSIM World 2011 was held in Virginia Beach, Virginia, October 11-14, 2011. The theme of the 2011 conference & expo was "Overcoming Critical Global Challenges with Modeling & Simulation". The conference program consisted of five technical tracks - Defense, Homeland Security & First Responders; Education; Health & Medicine; The Human Dimension; and Serious Games & Virtual Worlds.					
15. SUBJECT TERMS Modeling and simulation; Agent-based modeling; Game theory; Discrete event simulation; Cognitive modeling; Visualization					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	483	19b. TELEPHONE NUMBER (Include area code) (443) 757-5802